1989

Colonial Lawyer, Vol. 18, No. 1 (Spring 1989)

Editors of Colonial Lawyer

Repository Citation
https://scholarship.law.wm.edu/wmcl/30

Copyright © 1989 by the authors. This article is brought to you by the William & Mary Law School Scholarship Repository.
https://scholarship.law.wm.edu/wmcl
From the Nation's Oldest Law School

THE COLONIAL LAWYER:
A Journal of Virginia Law and Public Policy

VOLUME 18    SPRING 1989    NUMBER 1

Editor's Brief

Articles

Croson v. City of Richmond:
The Effect on Minority Business
Utilization Plans ...................................................... 1

Agricultural Preservation Techniques
in Virginia .......................................................... 11

Enforcement of the Safe Drinking
Water Act in Virginia ................................................. 33

State Control of Archaeological Resources
on Private Land .......................................................... 42

When the Zoning Power Comes Up Against
Freedom of Expression ................................................ 55

Negligent Infliction of Emotional Harm -
Let's Get Serious ....................................................... 75

The Case of "Intent" -
Should the Elements of Murder Be Expanded
in Virginia? .......................................................... 100

A STUDENT PUBLICATION OF
THE MARSHALL-WYTHE SCHOOL OF LAW
COLLEGE OF WILLIAM AND MARY
EDITOR’S BRIEF

As befits the spring volume of a publication, this issue of The Colonial Lawyer focuses to a large extent on land and the environment. Ms. Waymack’s article thoroughly discusses the rapid decline of usable farmland in Virginia and in other parts of the country, along with possible legislative remedies. Mr. Clark’s concern is with another resource derived from the land - our history, as discovered through the science of archaeology. This resource too is threatened. Ms. Munson concentrates on yet a third scarce and precious commodity, our drinking water.

Mr. Murphy’s article views land slightly differently. He recognizes the clash between certain First Amendment values and the police power of the state to effect zoning laws, and proposes a comprehensive analysis.

The recent Croson case concerning minority business set-aside programs originated in Richmond, so Mr. Dene’s and Mr. Franklin’s review of the law in this area is of regional as well as temporal significance.

Mr. Ehrler’s article assesses the history of awards for negligent infliction of emotional harm, and urges caution against what he perceives as the modern-day potential for over-compensation of the victim. He proposes concrete tests for determination of such awards, and distinguishes among types of harms.

In our final article, Mr. Thomas explores the statutory elements of murder in Virginia, and discusses whether different criteria for the “intent” component of the crime should be incorporated, as in other states.

I wish the best of luck to Thomas P. Sotelo and Lisa J. Entress, the incoming senior editor and managing editor, respectively. I hope you, the scholar and practitioner, continue to find the articles we at The Colonial Lawyer publish interesting and informative.

Felicia L. Silber
Senior Editor
CROSON V. CITY OF RICHMOND
THE EFFECT ON MINORITY BUSINESS UTILIZATION PLANS
Joseph F. Dene
James H. Franklin

INTRODUCTION

In 1983, the City Council of Richmond instituted the Minority Business Utilization Plan. The council heard testimony both in favor of and opposed to the plan. In opposing the plan, several construction trade associations stated their positions and fielded questions by council members on minority membership in their organizations. Generally, the groups had negligible or non-existent minority membership, but no evidence was presented to demonstrate that minorities had been discriminated against in attempting to enter such trade organizations. One group

1 RICHMOND, VA. CITY CODE (1985):
(a) All contractors awarded construction contracts by the city shall subcontract at least thirty (30) percent of the contract to minority business enterprises [MBEs]. Where the general contractor is a minority business this requirement shall be deemed to be met by the award....
The director of the department of general services shall be authorized to promulgate rules and regulations to implement the requirements of this division, which rules and regulations shall allow waivers in those individual situations where a contractor can prove to the satisfaction of the director that the requirements herein cannot be achieved.
(a) This division is remedial and is enacted for the purpose of promoting wider participation by minority business enterprises in the construction of public projects, either as general contractors or subcontractors.
(b) This division shall expire and terminate as of the last moment of June 30, 1988.

2 For example, last year we had the occasion to visit one company five times before we were even allowed to do business with them. On our fifth occasion going there, a vice president said to us, well, the National Football League did... not settle their dispute in one meeting, so you must learn to negotiate four, five, even a dozen times. We are marketing a standard product that is very similar to our competitors. I'm here speaking for contractors. True, I am not a contractor. I do not hire construction people. But I would not let this opportunity pass without saying that if you can give us access to the enterprise system, we want it. We're not asking to come in as second-rate service folk. We're asking to come in as first-rate service folk, but to allow us access. Thirty percent is not one hundred percent. There is still seventy percent left over. Adoption of Minority Business Utilization Plan: Hearing before the Richmond City Council, April 11, 1983 [hereinafter Hearing] (testimony of Ms. Cooper, at 17).

3 Id. (testimony of Mr. Watts, at 19; Mr. Beck, at 31; Mr. Murphy, at 35; Mr. Shuman, at 37).

4 Id.
testified that it actively sought to recruit minority members. Proponents of the plan were the only ones to present evidence of discrimination.

The J.A. Croson Company submitted the low bid on a project to install stainless steel urinals in the City Jail of Richmond. Croson experienced difficulty in obtaining minority suppliers to meet the city's thirty percent MBE requirement. The city denied Croson's request for a waiver, and decided to rebid the project. Croson then brought suit against the city.

In overturning the Minority Business Utilization Plan, the J.A. Croson v. City of Richmond decision has been regarded by the popular press as well as the dissent as the end to all affirmative action programs. This is an unfair characterization. Though the opinion holds race-based classifications to the strict scrutiny standard, it is still possible for a state or local government to responsibly implement a voluntary affirmative action program. Although the effect on existing affirmative action programs may be dramatic, Croson still allows for the development of effective affirmative action programs.

Strict scrutiny analysis requires that the government interest be compelling and that the means chosen are narrowly tailored to achieve that interest. Voluntary affirmative action plans by public employers must be narrowly tailored to remedy identified, prior governmental discrimination. If a public employer is using a statistical approach to prove prior discrimination, then the employer must use a restrictive view of the relevant labor market. A restrictive view would include only those minorities available to do a certain job, whereas a broad view would look to the

---

5 Id. at 38 (testimony of Mr. Shuman) ("Nobody that I know of, black, Puerto Rican or any minority, has ever been turned down. They're actually sought after to join, to become part of us.").

6 Id. at 41 (testimony of Mr. Marsh) ("There is some information, however, that I want to make sure we put in the record. I have been practicing law in this community since 1961, and I am familiar with the practices in the construction industry in this area, in the State, and around the nation. And I can say without equivocation, that the general conduct in the construction industry in this area, and the State and the nation, is one in which race discrimination and exclusion on the basis of race is widespread."). See also Hearings, at 42 (testimony of Mr. Deese).


8 Id.


10 Croson, 109 S. Ct. at 706, 739 (Brennan, J., dissenting).


13 Bakke, 438 U.S. at 291.

percentage of minorities in the general population. In Richmond, the percentage of available MBEs was unknown. The Court in *Croson* clarified the Supreme Court's standard in affirmative action cases by employing a restrictive view of the relevant labor market.

A public employer who chooses to use racial classification in a remedial program may only correct identifiable past wrongs; otherwise the scope of affirmative action would be unbounded. These wrongs may include passive participation by the city in the discriminatory process.\(^{15}\) The use of racial classifications should be limited for several reasons. First, often the burden of racial classifications falls on non-minorities who have never discriminated against minorities. Even benign racial classifications violate the anti-discrimination principle which holds that race is never to be used in the decision making process. Secondly, racial classifications are a dangerous tool; it is difficult to determine whether a plan is remedial or motivated by stereotypes and favoritism. Identifying past wrongs ensures a remedial nature. Finally, affirmative action should be limited because such race-based decision making may well reinforce the very stereotypes that lead to a need for remedial measures.

**THE TWO-PRONGED STRICT SCRUTINY TEST**

**FOR VOLUNTARY AFFIRMATIVE ACTION PROGRAMS**

Under strict scrutiny, the compelling interest requirement must be established by a substantial, factual showing of prior governmental discrimination.\(^{16}\) The means chosen must be "narrowly tailored" to remedy the identified discrimination.\(^{17}\) Justice O'Connor has stated that only a statutory violation, not a constitutional violation,\(^{18}\) is sufficient to establish prior discrimination.\(^{19}\) Further, it is sufficient to show that the government entity involved had a firm basis to believe that it had violated such a statute. Discrimination in fact is not required.\(^{20}\)

**CASE LAW**

In *Regents of the University of California v. Bakke*, a plurality of the Supreme Court held that a sixteen percent set-aside for minority medical school applicants was

\(^{15}\) *Croson*, 109 S. Ct. at 720.

\(^{16}\) *Bakke*, 438 U.S. at 309.


\(^{19}\) *Wygant*, 476 U.S. at 284 (O'Connor, J., concurring).

\(^{20}\) *Id.*
unlawful. Justice Powell used strict scrutiny because even a supposedly benign racial
classification is still suspect. Because the medical school made no showing of prior
discrimination, the school failed to establish the compelling interest required by strict
scrutiny review. Of the plurality in Bakke, only Justices Rehnquist and Stevens remain
on the Court. Justice White applied a mid-tier review stipulating that the means must
be reasonably related to an important governmental interest.

Fullilove v. Klutznick concerned a facial challenge to a federal program which
sought to remedy societal discrimination in the construction industry with the Public
Works Employment Act of 1977. This was a temporary, ten percent set-aside program
for minority business enterprises (MBEs). Congress was acting under the special
authority of the enforcement clause of the Fourteenth Amendment. Despite the lack
of showing that Congress ever discriminated against MBE's, another plurality of the
Supreme Court approved the set-aside plan in Fullilove, deferring to Congress as a co-
equal branch of government.

The district and appellate courts that first heard Croson deferred to the City
Council of Richmond as the Supreme Court had deferred to Congress in Fullilove. The
rationale for this deference is that Congress delegated its fact-finding authority to
localities under the 1964 Civil Rights Act. However, it does not necessarily follow
that congressional deference is imputed to localities merely because ease-of-
administration concerns warrant localized fact-finding. Congress merely delegated the
fact-finding process to the localities, which are better able to discover evidence of
localized discrimination. However, the localities must still do a responsible job, which
is measured by the standard commensurate with their resources.

21 438 U.S. 265 (1978) (four of the justices held the plan unlawful on Title VII
grounds, four justices held the plan lawful under Title VII, and Justice Powell held the
plan unconstitutional).

22 Justice White held that a private right of action under Title VI does not exist.
Bakke, 438 U.S. at 379 (White, J., concurring).

23 448 U.S. 448 (1980).

24 The societal discrimination was detailed in legislative fact-finding committee
reports. Id. at 459.


26 An MBE is defined as a company with fifty-one percent or greater minority
ownership. 42 U.S.C.S. Sec. 6705(f)(2).

27 U.S. Const. art. XIV, sec. 5.

28 Fullilove, 448 U.S. 448.

29 Fullilove, 448 U.S. at 492. See also Adelman, Voluntary Affirmative Action Plans
by Public Employers: The Disparity in Standards Between Title VII and the Equal
Chief Justice Rehnquist and Justice Stevens held that the Fullilove plan violated the Equal Protection Clause. This is significant because Congress was the actor, and challenge was only facial. Justices Brennan, Marshall, and Blackmun predictably upheld the MBE set-aside. Justice White also upheld the set-aside provision.

In Wygant v. Jackson Board of Education, a plurality struck down a race-based layoff system in a public school which maintained racial mix at the expense of the seniority of non-minority teachers. The school board relied on the student-teacher racial imbalance to justify their plan. However, the school board failed to show that it had ever discriminated against the teachers.

A plurality, including Chief Justice Burger and Justices Powell, Rehnquist, and O'Connor, required a finding of prior governmental discrimination to initiate a voluntary affirmative action set-aside plan. Justices Marshall, Brennan, and Blackmun believed that societal discrimination was a sufficient basis for race-conscious relief. Justice White voted against the plan in Wygant because he applied the anti-discrimination principle reasoning that race-based layoffs were never appropriate. Justice Stevens used a process approach; he offered no opinion on the issue of requiring prior governmental discrimination for race-based relief. Justice Stevens stated in United States v. Paradise that if the prior governmental discrimination is outrageous or gross, then a district court judge has broad authority to issue race-conscious relief.

To summarize, after Wygant, there are four members of the court who require a finding of prior governmental discrimination, three justices who do not require such a finding, one who allows affirmative action if the process is fair and one who has not addressed the issue.

In Johnson v. Transportation Agency, Santa Clara County, Cal., Justice Brennan wrote for a five-member majority that prior governmental discrimination is not required for the government agency to be able to implement remedial preferential relief. The majority held that remedying job classifications which have been traditionally segregated is a sufficiently important interest to warrant racial

30 Fullilove, 448 U.S. at 527, 552-53 (Stevens, J., dissenting).
31 Id. at 517 (Marshall, J., concurring).
32 Id. at 492.
34 Id. at 286 (Justice O'Connor's threshold was a "firm basis for belief that remedial action is required.").
36 Two of the four votes for a requirement of governmental discrimination have been replaced on the Court by Justices Scalia and Kennedy.
38 Justices Brennan, Marshall, Blackmun, Powell, and Stevens were the majority. Justice O'Connor concurred separately. Justice Powell is no longer on the Court.
classifications. Stevens concurred separately to emphasize that one may use race-based relief to remedy not only prior governmental discrimination, but also to increase services to ethnic groups, to ease racial tension, and to promote diversity.\(^39\)

In *Milliken*,\(^40\) the Supreme Court affirmed, without an opinion, the view that strict scrutiny is appropriate. Thus, a state must make a material, factual finding of prior discrimination before the state may use racial classifications.

**COMPELLING STATE INTEREST IN CROSON**

Chief Justice Rehnquist and Justices O'Connor, White, Kennedy, and Scalia require a showing of prior governmental discrimination.\(^41\) In *Croson*, Kennedy is squarely aligned with O'Connor and Rehnquist. Kennedy decided that strict scrutiny is the appropriate standard of review in lieu of an absolute ban on preferences to non-victims.\(^42\) White also has reaffirmed his opinion that a finding of prior governmental discrimination is required.\(^43\) Thus, after *Croson*, there are five Justices who require a finding of prior governmental discrimination.

It was the stated purpose of the Richmond set-aside plan to be remedial.\(^44\) The Court found that the city's reliance on Congressional findings, statements of council members, the lack of participation of minorities in trade associations, and the level of participation of MBE's in the public contracting industry\(^45\) was insufficient to show that Richmond had discriminated against MBEs. There was no direct evidence that the City of Richmond had ever discriminated against a minority contractor.\(^46\) Societal discrimination cannot be said to proximately cause the few number of MBE's because of the pervasive non-racial factors. Thus, a statistical disparity is insufficient to establish a remedial purpose in light of pervasive neutral explanations for the disparity.

---

\(^39\) *Johnson*, 480 U.S. at 647 (Justice Scalia favored a requirement of prior governmental discrimination. Justice White is no longer aligned with Brennan, Marshall and Blackmun, but rather now requires a material, factual finding of prior governmental discrimination. Rehnquist, O'Connor, Scalia and White are now opposed by Brennan, Marshall, Blackmun and Stevens. Justice Kennedy's position is critical.).

\(^40\) Michigan Road Builders Ass'n v. Milliken, 834 F.2d 583 (6th Cir. 1987).

\(^41\) *Croson*, 109 S. Ct. at 706.

\(^42\) *Id.* at 734.

\(^43\) *Id.* at 706.

\(^44\) *Richmond, Va. City Code Sec. 12-158(a)* (1985).

\(^45\) *See Hearing, supra* note 2, at 9.

\(^46\) *Croson*, 109 S. Ct. at 707.
The second prong of the strict scrutiny test asks if the classification is the least restrictive means available. Several components of the set-aside plan in *Croson* fail this test. The city council could have taken less restrictive steps to achieve the same goals. The plan could have been more narrowly tailored in several respects: if nonracial means had been utilized; if stricter measures had been used to define eligible minorities and only minorities from the Richmond area had been eligible; if the MBE participation percentage had conformed with previous Court opinions; and if a waiver provision had been incorporated to ensure that only MBEs from disadvantaged groups benefitted from the program. The city of Richmond seems to have been aware of the barriers MBEs face when attempting to enter the construction industry. Such barriers include inability to meet bonding requirements, lack of capital, unfamiliarity with bidding procedures and problems resulting from the lack of a track record. Instituting a race-neutral program to ease these requirements for fledgling construction contractors would have most likely resulted in a disparate impact benefiting black construction companies. Simplification of bidding procedures, relaxation of bonding requirements, and training and financial aid for disadvantaged entrepreneurs of all races would open the public contracting market to all those who have suffered the effects of past societal discrimination or neglect.

Thus, such a plan would enable MBEs to compete with established, non-minority companies and develop the capital, experience and business skills necessary to remain competitive. However, the City Council of Richmond does not appear to have considered such a race-neutral plan. Instead, the council opted for an affirmative action program, centered on explicit racial classifications. If a non-racial solution to the problem was ignored, the city of Richmond did not take the "least restrictive" route to achieve its aims. Had it attempted a non-racial approach and failed, then perhaps an explicit racial classification may have been appropriate.

The city of Richmond's plan specified minority groups that were eligible to participate in the set-aside program. The list was identical to the one used by Congress in the Public Works Employment Act of 1977, the federal set-aside legislation validated in *Fullilove*. Congress was creating a program which was national in scope; it appropriately included various minority groups such as Hispanics, Aleut, and Eskimos. In Richmond, though, the Eskimo and Aleutian populations are negligible, if they exist at all; it is quite unlikely that the city government or the local construction

---

47 *Fullilove*, 448 U.S. at 480.

48 *Croson*, 109 S. Ct. at 708.

49 Id. at 729.

50 See Hearing, supra note 2, at 1.

51 *Fullilove*, 448 U.S. at 455 (The list included citizens who are: "... Negroes, Spanish-speaking, Oriental, Indians, Eskimos, and Aleuts.").
trade associations discriminated against either one of these groups. By including groups that were never victimized by the local government or even Richmond society in general, the ordinance's remedial purpose is impugned.\textsuperscript{52} This gives an unfair preference to those groups based solely on race.

Additionally, the Richmond plan lacks any type of geographical limitation. Thus, minorities from anywhere else in the country receive an advantage over local non-minority contractors, even though they had never been discriminated against in the Richmond area. Over-inclusion of groups that are not the victims of discrimination negates the remedial character of the set-aside.

The third reason that the plan failed the "least restrictive means" criterion is that the ordinance set a thirty percent requirement for minority participation in city construction contracts.\textsuperscript{53} The plan roughly split the difference between the percentage of blacks in the city population, about fifty percent, and the percentage of MBEs who received city contracts, less than one percent, to achieve a figure of thirty percent. This is the formula used by Congress in the Public Works Employment Act of 1977.\textsuperscript{54} Despite the holding in \textit{Fullilove}, the Court has said that this is not the proper formula to determine a percentage set-aside: "But where special qualifications are necessary, the relevant statistical pool for purposes of demonstrating discriminatory exclusion must be the number of minorities qualified to undertake the particular task."\textsuperscript{55} Because construction contractors possess special qualifications, the total percentage of minorities in the general population seems to be irrelevant in determining percentages for set-aside programs.\textsuperscript{56}

Though the city council followed the congressional set-aside plan validated in \textit{Fullilove} in many respects, it did not do so completely. The city provided a waiver in the event minority contractors were unavailable to fulfill a city contract.\textsuperscript{57} Unlike Congress' plan however, "...there is no inquiry into whether or not the particular MBE seeking a racial preference has suffered from the effects of past discrimination by the

\textsuperscript{52} \textit{Croson}, 109 S. Ct. at 728.

\textsuperscript{53} \textit{Richmond, Va. City Code Sec. 12-156(a) (1985)}.

\textsuperscript{54} \textit{Fullilove}, 448 U.S. at 465.


\textsuperscript{56} The dissent makes a very compelling argument to the contrary. If Blacks have been excluded from the construction industry then the applicable figure to examine in calculating a quota is the percentage of blacks in the general population. This assumes that if there had not been discrimination, then blacks would have a market share approximately equal to their percentage in the general population. \textit{Croson}, 109 S. Ct. at 746.

\textsuperscript{57} \textit{Richmond, Va. City Code Sec. 12-157} (1985).
Such a flexible waiver provision would have safeguarded the remedial goals of the plan and innocent parties would not have been required to shoulder unnecessary burdens. Furthermore, such a waiver would have minimized the over-inclusion of groups that were never the victims of discrimination in Richmond.

Finally, the Richmond ordinance had a significantly longer duration, five years, than the plan in Fullilove. Though the Court has not set a specific limit on the duration of affirmative action programs, the more brief the program the less restrictive it appears. However, the duration of set-aside plans does not appear to be a major component in the least restrictive means segment of the strict scrutiny test.

AN ALTERNATIVE MBE UTILIZATION PLAN

A state or local government should first consider the use of facially neutral means. A facially neutral law, on the Croson facts, could be designed to have a disparate impact on minorities and thus would address the bonding problem, working capital deficiencies, unfamiliarity with bidding procedures, and the lack of a track record often facing new minority firms. A facially neutral approach additionally has the advantage of not being over-inclusive. Furthermore, because no racial classification is used, the plan does not need to pass strict scrutiny tests and the legislation is less likely to be challenged.

To increase the level of MBE participation in municipal contracts, a city should determine if there is a firm basis for belief that there was local governmental discrimination. The Court has held that if there has been governmental discrimination, then race specific measures are permissible. If no governmental discrimination exists, then nonracial measures must be used.

The first step to ameliorating the discrimination is identification of the wrong that is to be remedied. Then remedial measures should be narrowly tailored so as not to unnecessarily trammel the interests of non-minorities. In a set-aside program with goals similar to those in Croson, the plan should be limited to those groups who have been victims of discrimination.

The percentages in a model plan would reflect the percentage of available MBEs in the relevant market and not the percentage of minorities in the general population. The number of available MBEs should increase as new MBEs enter the market to compete with existing MBEs. If the MBEs are not working at full capacity, there will always be more available MBEs than participating MBEs. As the percentage of available

58 Croson, 109 S. Ct. at 729.
60 See Croson, 109 S. Ct. at 728.
61 See Bakke, 438 U.S. at 265; Milliken, 834 F.2d at 583.
62 See supra, note 63.
MBEs increases, their statutory share of the market could rise with them. This increase could continue to a reasonable level such as half way between the current level of MBE participation and the percentage of minorities in the population, as long as the level was predetermined. The minority percentage in the population would provide an upper limit. After this maximum point has been reached or the statutory time limit has expired, the statutory percentages would symmetrically decrease. The MBEs would gradually lose their government preference and thus compete in the general construction market.

The model plan would also be limited to a geographic area equivalent to the relevant market. This geographic limitation serves a fairness function; it is patently unfair for remote, non-victims to have a preference over non-minority, non-discriminatory contractors.

CONCLUSION

The Supreme Court has established that strict scrutiny is the appropriate standard of review in affirmative action cases. Even though set-aside programs are subject to strict review, they are still viable. By responsibly establishing a firm basis for belief of prior governmental discrimination and by narrowly tailoring the steps to remedy that discrimination, a city or state government may still implement an effective affirmative action program.

---

63 The model plan should have a time limit at which the predetermined point will be considered to have been reached, whereupon the set-aside is decreased.

64 "Symmetrically" means that the percentage will decrease by the same amounts and over the same time periods as it was increased.
AGRICULTURAL PRESERVATION TECHNIQUES IN VIRGINIA
Jacqueline Waymack

INTRODUCTION

Large amounts of agricultural land within the United States have been irreversibly converted to residential, transportation, commercial and other urban uses within the last two decades. The total number of acres devoted to farming has declined nationwide in the past twenty years.\(^1\) Confronted with statistics showing a dramatic shift in land use, many states implemented legislation aimed at preventing the loss of farmland. In the early 1980's, the U.S. Department of Agriculture conducted a large study of farmland conversion and the programs that various states had adopted to combat the loss of such lands. The National Agricultural Lands Study concluded that agricultural lands were disappearing at a dramatic rate across the nation and proposed that states adopt comprehensive growth management programs to halt the rapid loss of economically and environmentally important lands.\(^2\)

Land use patterns in Virginia have followed the national trend of decreasing agricultural land use and increasing residential and urban land uses.\(^3\) The amount of land devoted to agricultural production has steadily declined since 1960. In 1960 Virginia had 13.5 million acres of farmland; in 1970, 11.4 million acres; in 1975, 10.1 million acres; in 1980, 9.8 million acres; in 1988 9.6 million acres.\(^4\) Conversion of agricultural land in the eastern part of Virginia has occurred more rapidly than in other regions and accounts for most of the shift in land use within the state.\(^5\)

Many factors contribute to the conversion of farmland. Urban growth pressure, farm profitability, land value, taxes, government programs, regulations and incentives, community expectations and personal decisions about work and retirement all affect the conversion process.\(^6\) As urban growth spreads into the countryside, land values increase

---


\(^2\) NALS, supra note 1. The study projected that a 60% to 85% increase in demand for U.S. agricultural goods would occur in the next 20 years.

\(^3\) U.S. DEPARTMENT OF AGRICULTURE, SOIL CONSERVATION SERVICE, VIRGINIA NATIONAL RESOURCES INVENTORY DATA 7 (1982) [hereinafter VNRID].


\(^5\) VNRID, supra note 3, at 7.

\(^6\) NALS, supra note 1, at 16.
and investors begin buying land for development. When scattered development occurs, farming becomes more expensive and farm expansion more difficult. Public services such as water and sewer facilities extend into an area to meet the needs of new residents and enable the area to support more development. Public agencies take land by eminent domain for highways, parks, powerlines and other facilities necessary to support the influx of people. Real estate taxes increase because of these new public services. Consequently, some farmers sell their lands as a result of development pressure; others join the bandwagon in the belief that agriculture in their area is doomed.\(^{7}\)

A majority of states have implemented some type of farmland preservation program in order to protect important farmlands from development. These programs often include a variety of statutes aimed at making farm areas less attractive to developers, easing the burdens of development for farmers, and preventing farmland from being converted to non-agricultural use.\(^{8}\) This article examines the agricultural preservation efforts in Virginia. It analyzes five pieces of legislation, comments on their effectiveness and suggests some changes that would better protect important agricultural land from conversion.

FARMLAND PRESERVATION IN VIRGINIA

By 1981, Virginia had implemented several pieces of legislation aimed at preserving agricultural land within the state. These include differential tax assessment for farmlands, agricultural districting, right to farm legislation, agricultural zoning, and a farmland preservation act.\(^{9}\) Overall, these measures reflect a fairly conservative approach to farm preservation. Each of these acts has a policy provision that affirms the state's commitment to the preservation of farmland as an economic and environmental resource. Nevertheless, the main effect of these acts is to ease the burden of encroaching development on farmers while providing some relief from governmental interference to one sector of the state's economy.

These programs do provide some incentives for farmers to keep farming, but they do not prevent farmland from being converted to more intensive uses. The voluntary nature of the tax relief and agricultural districting acts limits their potential to prevent conversion of farmland. The ambiguities and loopholes present in the right-to-farm legislation and the restricted scope of the preservation of prime agricultural act limit the value of these two acts as tools for farmland preservation. Agricultural zoning provides an effective tool to restrict the amount of farmland eligible for conversion, but its implementation remains in the hands of local governing bodies that may have

---

\(^{7}\) *Id.* at 35-37.

\(^{8}\) *Id.* at 37-38.

\(^{9}\) The tax assessment statute was enacted in 1971; the agricultural and forestland districting act in 1977; the right to farm act in 1981; the zoning statute in 1950; and the preservation of prime agricultural land act in 1981.
little interest in preservation. However, they do form a good foundation on which to build.

Differential Tax Assessment

Designed to correct the increase in land value that accompanies encroaching development, differential tax assessment permits the farmer to have his land appraised at its agricultural use value instead of its fair market value. It is the most common form of tax relief that states have adopted to further the goal of farmland preservation.\(^\text{10}\) There are three different types of differential tax assessments: preferential, restrictive agreement and deferred.\(^\text{11}\) Some states have adopted preferential tax assessment programs in which eligible land is assessed at its use value and no penalty attaches for conversion of the land to a non-eligible use.\(^\text{12}\) Few states have adopted restrictive agreement programs. In order to obtain differential assessment under this type of tax relief program, landowners must enter into a contract prohibiting the conversion of their land.\(^\text{13}\) A majority of states, including Virginia, have deferred tax assessment laws. Under this type of statute, penalties attach to the withdrawal of land from the use value assessment program.\(^\text{14}\)

Many localities in Virginia have adopted a use value taxation scheme pursuant to the deferral tax assessment statute.\(^\text{15}\) Participation by both localities and individuals is voluntary. Localities must pass ordinances adopting the tax relief measure and individual landowners must apply to the local assessing officer in order to receive any tax relief.\(^\text{16}\) A farmowner can receive special assessment of his land and the structures present on the property. Qualifying structures include any structure, except farm houses, which is used in connection with the land's special use.\(^\text{17}\) Qualifying land must meet several criteria before the special tax assessment can apply.

First, the land must be devoted to either agricultural, forestal or open space use. The determination of the land's use is the responsibility of the Director of the

\(^{10}\) K. MEYER, D. PEDERSON, N. THORSON & J. DAVIDSON, AGRICULTURAL LAW 853 (1985) [hereinafter TEXT]. Every state except Kansas has some program for reducing the tax burden on farmers.

\(^{11}\) NALS, supra note 1, at 18.

\(^{12}\) See NALS, supra note 1, at 19; TEXT, supra note 10, at 853 n.1. Seventeen states have preferential tax assessment, including Florida, Iowa and the Dakotas.

\(^{13}\) See NALS, supra note 1, at 18; TEXT, supra note 10, at 853 n.1. California and New Hampshire have restrictive agreements.

\(^{14}\) TEXT, supra note 10, at 853 n.1.

\(^{15}\) Statistics were unavailable for the number of acres in the states which receive use value taxation. A majority of, but not all, counties have land use taxation.


\(^{17}\) Id. at Sec. 58.1-3236(B)(1984).
Department of Conservation and Historical Resources and the Commissioner of Agriculture and Consumer Services. The special tax assessment is also available for lands which are part of an agricultural or forest district. In 1987, the legislature amended the tax code to provide that, regardless of any deferred tax assessment ordinances, local governments can avoid the application of use value assessment by designating land in certain areas as ineligible for special assessment. Thus, a locality interested in developing a certain area can tax agricultural, forest and open space land in that area at fair market value. This frustrates the purpose of deferred taxation (preventing conversion, or at least easing the burdens of development) by allowing localities to deny tax relief whenever they desire development of an area. While this provision does give more flexibility to the locality for guiding growth, it certainly weakens the act as a land preservation tool.

In addition, the land must be a certain size in order to qualify for special assessment. Land devoted to agricultural, horticultural or open space use must consist of at least five acres; land in forestal use must cover at least twenty acres. Except for properties divided by public right of way, land must be in contiguous parcels in order to satisfy these minimum acre requirements. Because the number of small farms (1-180 acres) is growing in Virginia, such a minimum acre requirement allows more farmers to benefit from the special tax assessment.

Virginia uses the capitalization of income method to determine the use value of qualified land. Each year the State Land Evaluation Advisory Committee [SLEAC] establishes a range of suggestive values for land in each soil classification based on the productive earning power of the property as it is used for agricultural, horticultural or open space purposes. SLEAC publishes these value ranges for the counties and cities which use them as a guide to determine the actual value of land within their jurisdiction. To define the productive earning power, the Committee either looks to the capitalization of the warranted cash rents of the specific property or the incomes of similar properties in the locality.

The deferred tax assessment statute provides for penalties for misstatements made in applications, changes in the use of the land, failures to report such changes and for applying for and receiving a rezoning of the land to a more intensive, nonqualifying


20 *Id.* at Sec. 58.1-3231.

21 *Id.* at Sec. 58.1-3233(2) (1984 & Supp. 1988). Open space land in cities must be at least two acres.

22 *Id.* at Sec. 58.1-3233(2) (Supp. 1988). Recorded subdivision lots owned by the same person cannot be included to meet the minimum acre requirement.

23 1982 CENSUS, *supra* note 1, at Part 46 at VIII. From 1978-1982, the number of farms that have 1-180 acres of land have increased by 2289 farms.

use. Persons who either make material misstatements in their applications or who fail to report changes in land use are responsible for any taxes which would have been levied otherwise. If a landowner makes a material misstatement with an intent to defraud, then she must pay an additional penalty of 100% of the unpaid taxes. While the statute lists some things that constitute material misstatements, it does not provide any examples of reportable changes in use.

A penalty of additional taxes attaches if a landowner either changes the use of her land so that it no longer qualifies for special assessment or requests and succeeds in having that land rezoned to a more extensive use which is not compatible with an agricultural use. This additional tax is known as a roll-back tax and is equal to the difference between the tax which the landowner would have paid had the land not been under the special assessment program during the past five years plus simple interest. An owner must report a change in use or a rezoning of her land within 60 days after it occurs to the local commissioner of the revenue or the assessing officer.

Once the commissioner has determined the amount of the roll-back tax, the landowner has 30 days to pay or she is assessed an additional penalty. If rezoned land is later zoned for agricultural use, it is not eligible for special land use assessment for three years. However, if the land is taken by eminent domain or a change in ownership occurs, no such penalty accrues. Thus, a farmer can sell her land to someone else who then can develop that land without suffering the withdrawal penalty. Those landowners with good tax advice avoid the penalties for developing their lands.

While tax relief for land devoted to agricultural uses is widely viewed as an important method for preserving farmland, it does not prevent farmland from being converted to other uses. The underlying assumption of differential tax assessment legislation, that tax burdens have a great influence over whether a parcel of land is converted to nonagricultural use, is faulty. Because a multitude of factors influence the conversion of land to nonagricultural uses, decreasing the tax burden for farmers may have little impact on whether land remains in agricultural use. Because of the voluntary nature of the land use taxation program and the weak penalties for

---

25 Id. at Sec. 58.1-3238 (1984).

26 Id. at Sec. 58.1-3237(A)(D) (Supp. 1988).

27 Id. at Sec. 58.1-3237(B) (Supp. 1988).

28 Id. at Sec. 58.1-3237(C) (Supp. 1988).

29 Id. at Secs. 58.1-3237(C), 58.1-3234 (Supp. 1988).


31 Dunford, A Survey of Property Tax Relief Programs for the Retention of Agricultural and Open Space Lands, 15 GONZ. L. REV. 675, 689 (1980).

32 Id. at 689-90.
withdrawal, the act affects the preservation of land only to the extent that farmers find it economically worthwhile to participate in it. Thus, even though the special tax assessment in Virginia may postpone conversion of lands to some future time, it is not a preservation measure; instead it operates to offset the tax burden caused by approaching development in order to keep farming economically viable for those farmers who want to farm.

The differential tax assessment actually shifts the tax burden to other groups within the taxing jurisdiction. Local governments may increase the taxes of nonagricultural taxpayers in order to compensate for the loss of income from farmers. Such tax shifting may erode the tax base of rural areas, adversely affecting the quality and availability of services within those areas. This loss of tax base problem could affect the willingness of localities to adopt land use assessment ordinances. Indeed, poorer counties, such as Charles City, have no special assessment for farmland.

AGRICULTURAL DISTRICTING: THE AGRICULTURE AND FOREST ACT

Virginia has also adopted agricultural districting in order to keep land in agricultural use. Districting emphasizes long term preservation of farmland through voluntary compliance and local initiative. It provides an incentive for farmers to continue farming by conferring benefits to those who voluntarily place their land in special districts. Those lands included in districts are taxed at use value and protected from government activities aimed at development.

Currently, Virginia has 547,095.77 acres in 175 agricultural districts located in twenty-two counties and one town across the state. The acres included in the

33 Id. at 695.
34 NALS, supra note 1, at 37-38.
35 Dunford, supra note 31, at 692.
36 Id. at 693.
38 Id.
39 Statistics provided by the Virginia Department of Agriculture & Consumer Services, Richmond, Va. Accomack Co. has the most number of acres in districts--83,702.43 acres in 22 districts; Fauquier Co. has the second most--81,988.20 acres in 9 districts; Loudoun Co. has the third most--66,819.91 in 17 districts. Fairfax Co. has 2 districts with 1,261.36 acres and 23 local districts with 2,646.75 acres. Blacksburg is the only town that has any districts; it has 2,529.44 acres in 1 district. Remaining counties: Albemarle (44,962.52 acres in 13 districts), Clarke (24,959.40 acres in 2 districts), Culpeper (45,736 acres in 13 districts), Frederick (15,013.58 acres in 1 district), Green (23,597.43 acres in 7 districts), Hanover (14,301.86 acres in 8 districts), Isle of Wight (8,191 acres in 2 districts), James City (18,209.30 acres in 12 districts), King William (3,729.95 acres in 5 districts), Madison (607.07 acres in 1 district), Montgomery (47,487.91 acres in 11 districts), New Kent (15,241.96 acres in 8 districts), Orange (668 acres in 1 district), Prince William (3,466.83 acres in 2 districts), Rappahannock (16,841.12 acres in 9 districts), Shenandoah (9,702.75 acres in 4 district), Tazewell (7,362 acres in 1
Agricultural districts comprise only 0.06% of the state's farmland. Counties in the northern-northwestern region of the state have the most districts, while counties in the southern portion have few, if any. In the Tidewater region, only James City and Isle of Wight have land in districts and both of these are on the outskirts of the Hampton Roads sprawl. In the Richmond-Petersburg area, only Hanover and New Kent Counties have land in districts and these two remain primarily agricultural. In southwest Virginia, Tazwell and Montgomery Counties have districts; on the eastern shore, Accomack County has over 83,700 acres in agricultural districts. 40

The state purpose in enacting the districting statute is the protection and enhancement of agricultural and forest land as a viable segment of the state's economy and a valuable economic and environmental resource. 41 The Act declares that it is a state policy to protect and conserve these lands as important "ecological resources which provide essential open spaces for clean air sheds, watershed protection, wildlife habitat, as well as for aesthetic purposes." 42 However, the provisions within the Act are directed toward the economic aspect of land preservation and do little to implement the policy of ecological protection. Like the tax relief statute, this act focuses on offsetting the burdens caused by encroaching development rather than preventing the conversion of farmland.

The act applies only to land which has been, is currently or can be retained for agricultural or forestland production. 43 While land which is not currently devoted to production may be included in a district, it may not receive some of the benefits relating to the restriction of governmental activity. Likewise, land devoted to nonagricultural use does not receive any benefits under the act. The statute does not offer any preservation methods to counties that want to preserve open space lands from private development or government interference.

Districts can only exist in the localities which have adopted an ordinance permitting their creation. 44 The absence of such ordinances may account for the lack of districts in certain areas, especially those which faced strong development pressures in the early 1980's. Because some local interest in the districting program must exist before a governing body will adopt it, farmers and members of local governing bodies in areas of high farm conversion rates probably were not interested in establishing districts when the state adopted the districting act. Likewise, rural areas not confronted with approaching development or those wishing to attract development to boost the local economy may have had little interest in implementing such a program.

40 Id.
42 Id.
44 Id. at Sec. 15.1-1511 (1981 & Supp. 1988).
Once a locality does pass such an ordinance, landowners can submit applications to the local governing body requesting to create a district.\(^\text{45}\) The locality provides the application forms and may charge a fee, which cannot exceed the cost of processing or $300, whichever is less.\(^\text{46}\) Thus the base of the districting program rests on the willingness of farmers with significant landholdings to request that their lands form a district.

The Virginia statute requires that each district contain at least 200 acres of land in a contiguous parcel.\(^\text{47}\) The land may be owned by one or more persons.\(^\text{48}\) Land not contiguous with the 200 acre core may be included in the district if it lies within one mile of the district's boundary or if it borders on non-core land within the district.\(^\text{49}\) A district may extend between two localities provided that both governing bodies grant approval of it.\(^\text{50}\) The 200 acre limit may not be sufficient for the needs of Virginia farmers. The average size farm in Virginia is 196 acres and the number of smaller farms (1-180 acres) is growing.\(^\text{51}\) A lower limit would provide more protection, especially for counties where development has swallowed most of the farmland.

Once the local governing body receives a districting application, the planning committee provides public notice, holds hearings on the matter and makes recommendations to the local governing body.\(^\text{52}\) The land may be evaluated through the Virginia Land Evaluation and Site Assessment System [LESA] or a local LESA system if one exists.\(^\text{53}\) The statute lists several factors which the committee should consider when evaluating land. These include the agricultural or forestal significance of land to be included in and adjacent to the district, the presence of any such significant land which is not in agricultural or forestal use, the extent of other uses for land in and around the district, local development patterns, needs and zoning regulations and the environmental benefits of creating a district.\(^\text{54}\) Once the local governing body receives

\(\text{\textsuperscript{45} Id. at Sec. 15.1-1511.}\)

\(\text{\textsuperscript{46} Id. at Sec. 15.1-1509 (1981 & Supp. 1988). A sample application is provided in the statute.}\)

\(\text{\textsuperscript{47} Id. at Sec. 15.1-1511(D) (Supp. 1988). Earlier versions of the statute required a 500 acre minimum.}\)

\(\text{\textsuperscript{48} Id. at Sec. 15.1-1511(A) (Supp. 1988). Earlier versions of the act had a 3500 acre limit for any one owner.}\)

\(\text{\textsuperscript{49} Id.}\)

\(\text{\textsuperscript{50} Id.}\)

\(\text{\textsuperscript{51} 1982 CENSUS, supra note 1, at Part 46 at VIII. Statistics on the average farm size provided by the Va. Agric. Statistics Service.}\)

\(\text{\textsuperscript{52} VA. CODE ANN. Sec. 15.1-1511(B)(1) (Supp. 1988).}\)

\(\text{\textsuperscript{53} Id. at Sec. 15.1-1511(C) (Supp. 1988).}\)

\(\text{\textsuperscript{54} Id. at Sec. 15.1-1511(C).}\)
the committee’s final report, it holds a public hearing and may pass an ordinance to create a district. The locality may condition the creation of a district on landowners’ obtaining governmental approval before land in the district can be used more intensively or for a more intensive nonagricultural use. After the creation of a district, the local governing may adopt incentive programs which promote certain land use and conservation in a district.

Districts may be reviewed no less than four years and no more than ten years after their establishment. If lands are withdrawn from the district so that it no longer has a 200 acre core or if annexation of the land occurs, the districting ordinance remains in effect until the date set for review. Land may be added to a district by following the creation procedure.

Landowners may remove land from the district by filing a written notice to the local governing body. If the notice is filed at the time of the district review, the landowner must submit written notice to the governing body before it has acted with regard to that review. If a landowner files a notice at any other time, the planning committee must make recommendations regarding the request and public hearings must be held on the matter. If the governing body rejects the landowner’s request, the landowner has a right of appeal de novo to the circuit court. If the governing body approves the withdrawal, the land is subject to roll-back taxes and to the government actions that were restricted while the land was in the district. If a landowner dies, his heirs can withdraw the land by filing written notice to the governing body within two years after the owner’s death. Thus, withdrawal from the program is quite easy.

The creation of agricultural or forestal districts confers two general benefits to landowners: land use tax assessment and the restriction of government activities. Land within a district automatically qualifies for use-value tax assessment if it meets the requirements of the deferred assessment statute. Land within a district must be devoted to an agricultural or forestal use in order for the special assessment to apply. Unfortunately, the same problems of ineffectiveness and tax shifting arise in this

55 Id. at Sec. 15.1-1511(D).
56 Id. at Sec. 15.1-1511(F) (Supp. 1988).
57 Id. at Secs. 15.1-1511(D), 15.1-1513(F) (Supp. 1988).
58 Id. at Sec. 15.1-1511(E) (Supp. 1988).
59 Id. at Sec. 15.1-1511(F).
60 Id. at Sec. 15.1-1513(A) (Supp. 1988).
61 Id.
63 Id. at Sec. 15.1-1513(D) (Supp. 1988).
64 Id. at Sec. 15.1-1512(A) (1981 & Supp. 1988).
portion of the districting act that appears in the tax relief statute. Thus, one of the two major incentives for farmers to place their lands in districts does not effect the Act's main goal of preservation.

The more important benefit derived from placing land in a district is the restriction on governmental activity. The statute ostensibly limits actions of both local and state governments. A locality cannot enact ordinances which would unreasonably restrict or regulate farming, farm structures or forestry practices in a district unless such regulation has a direct relationship to public health and safety. However, because the Act only prohibits a locality from unreasonable regulation, it adds very little to existing limitations on governmental power. Indeed, the limitation merely provides a functional definition of the locality's police power.

The Act also provides that local water, sewer, electricity or non-farm/non-forest drainage districts cannot impose benefit assessments or special tax levies on land which is primarily devoted to agricultural or forestland production within a district. This limitation specifically focuses on inhibiting the urbanization process by relieving the developmental pressures produced when localities finance public improvements in their areas through special assessments. However, by including comprehensive planning at the district formation stage, the act has provided some leeway for farmers who wish to obtain an increase in public services in order to intensify their farm operations. Additionally, localities can maintain special assessments for one-half acre lots surrounding dwellings or non-farm structures within a district. Before the exemption from special levies attaches, land must meet a use requirement even though it is already included in a district. Such a requirement is consistent with the statute's policy of preservation of land as an economic resource, but does not encourage preservation of land for environmental purposes.

The statute restricts the state's exercise of eminent domain over lands within districts and establishes their maintenance as a policy of state agencies. If any state agency intends to take land within a district by eminent domain, it must notify the

---

65 Id. at Sec. 15.1-1512(B) (1981 & Supp. 1988).
66 Meyers, supra note 38, at 643.
67 Id.
69 Meyers, supra note 38, at 643.
70 Id.
72 Id. at Secs. 15.1-1512(E), 15.1-1508(5) (1981 & Supp. 1988). "Agriculturally significant land" is land which either has produced agricultural or forest products or is currently producing or considered important farmland.
73 Id. at Sec. 15.1-1512(D) (1981 & Supp. 1988).
local governing body. The locality then has 30 days to review the state's proposed action. If the locality determines that the action is contrary to state or local policy, it can order the agency to wait an additional 60 days before taking any action. During that period, the locality holds public hearings and issues a final order concerning the proposed action. If that final order decrees that the agency may not act because such actions would adversely affect state or local policy, the agency may appeal the order in the circuit court having jurisdiction over the district.74 These express procedural limitations on the state's exercise of eminent domain appear to be a significant tool for preserving farmland and guiding growth in rural areas.75

While the eminent domain protection for land in a district effectively prevents land from converting to certain intensified uses, it applies only in the limited context of state action so that it only protects against conversion which a landowner probably would oppose anyway. While this provides a good benefit for landowners who do not want their lands to become highways, it does not address the loss of farmland caused by economic development. The act does not offer farmers faced with increased development pressures any incentives, other than the threat of roll-back taxes, to keep their lands in agricultural uses. Instead of actually preventing conversion of farmland, forestland or open space land, the act merely provides a few incentives to the state's agricultural economy mainly by limiting state actions which interfere with lands devoted to production.

The state's policy provision is unlikely to grant farmers any additional rights or to protect against farmland conversion.76 While state agencies are required to encourage the maintenance of farming and forestry within districts,77 they are not prohibited from adversely affecting the land within a district.78 Nevertheless, the existence of state policy to maintain farming in districts could be significant during a judicial review of an agency's actions.79

RIGHT-TO-FARM ACT

Virginia's right to farm act insulates farmers from nuisance liability in certain situations, thus, protecting some agricultural land from conversion.80 It focuses on alleviating the problem of land-use conflict that arises when development borders on

74 Id.
75 Meyers, supra note 38, at 644.
76 Id.
78 Meyers, supra note 38, at 644.
79 Id.
agricultural production areas. When more and more people move into farm areas, residential uses of land begin to conflict with agricultural operations. The new residents object to the noise, smells, pesticides, dust and other by-products associated with farms. These people express their discontent in the form of nuisance suits or by influencing the local governing body to pass an ordinance which declares the farm operation a nuisance. Indeed, nuisance suits have been a problem for Virginia farmers. Defending against nuisance actions and challenging ordinances is expensive for farmers and if they lose, they may have to halt their farm operations. The Act can protect farmers from becoming victims of urban sprawl only by barring nuisance claims and voiding local ordinances in this situation.

The Act does not protect farmers either when their rural neighbors bring nuisance actions or when the conflicting residential use existed prior to the farm operation. It only applies when new residents arrive in the area and initiate actions against pre-existing operations. Thus, the statute codifies the "coming to the nuisance" defense. The Act states that a court cannot declare an agricultural operation a nuisance if it has been operating for one year prior to a change in condition in that locality. Little doubt exists that the legislature intended the change-in-condition clause to refer to urban sprawl; the act's policy declaration specifically identifies the extension of nonagricultural land uses into agricultural area as the cause of farm nuisance suits. In an interpretation of a similar right-to-farm act, the Georgia Supreme Court concluded that "changed conditions in...the locality" refer[red]...solely to the extension of nonagricultural land uses.... Thus, in farm nuisance actions, the date of the change in condition becomes extremely important; however, determining the date when that change occurred is difficult.

The statute modifies the "coming to the nuisance" defense by creating a statute of limitations. The act restricts new neighbors from bringing nuisance actions and localities from adopting nuisance ordinances if the farm operation in question has been functioning for more than one year. Such a limitation discourages farmers concerned about potential nuisance actions from establishing new operations, especially feedlots,

---

81 VA. CODE ANN. Sec. 3.1-22.28 (1983). Agricultural operations are defined as any operation producing crops, livestock, nursery, floral or forest products for sale. Id. at Sec. 3.1-22.29(B).

82 NALS, supra note 1, at 21.

83 No reported cases under the right to farm act could be found for Virginia. However, officials at the Va. Agric. Statistics Service were aware of nuisance suits being brought against farmers in the state both prior to and after the passage of the right-to-farm act.

84 Grossman & Fisher, supra note 80, at 122.

85 VA. CODE ANN. Sec. 3.1-22.29(A) (1983).


in areas threatened by development. A developer acquiring land near a young farm operation could easily use nuisance actions to either force the operation to close or contribute to its demise. Conversion of the farmland to nonagricultural uses could easily follow. Thus, the one-year limit favors the expansion of nonagricultural uses over agricultural uses, especially in localities where the two are just beginning to conflict. While the year limit would protect new residents from objectionable activities of new farm operations, it does not further the state's policy of protecting, conserving and encouraging the development of agricultural land. Indeed, the discouragement of agricultural expansion is inconsistent with farmland preservation.

The statutory prohibition against declaring a farm operation a nuisance does not apply to situations where the alleged nuisance resulted from negligence, or improper operation. The act does not define either term. While a court could easily determine the meaning of negligence, it might have difficulties with a vague term like "improper". Some states' right-to-farm statutes include definitions of both negligence and improper standards. States like New Hampshire and Idaho define both terms with respect to compliance with local, state and federal laws and regulations. Including a provision which defines improper operation as a violation of federal, state, or local law or regulation would certainly clarify the Virginia statute. Such a liability standard for farm operations would also accommodate the dual state policies of farmland preservation and environmental protection. Indeed, allowing farmers who pollute or do not exercise due care for the rights of others to escape nuisance liability does not further the act's policy.

The Act also denies protection from nuisance actions when the farm operation has significantly changed. Expansion or other change in the farm operation could remove the Act's nuisance protection completely. While this provision may protect new residents from offensive activities, it also places the farmer in the dilemma of choosing between more efficient farming methods and protection from nuisance liability. Such a situation clearly does not further the act's policy. However, the interpretation of "significant change" determines the extent of the Act's protection.

"Significant change" is a very ambiguous phrase with a wide range of interpretation. If "significant" is construed strictly, the act could penalize farmers for diversifying or investing in a more profitable crop or more efficient farming method. Such a result frustrates the policy behind the Act. If "significant" is interpreted very broadly, its inclusion in the Act could become meaningless as almost nothing would be

---

88 Id. at Sec. 3.1-22.29(A).
89 Grossman & Fisher, supra note 80, at 132.
90 Id.
91 Id.
93 Grossman & Fisher, supra note 80, at 128.
considered significant. Such an interpretation is very unlikely. Courts addressing this unresolved issue would probably decide the existence of a significant change on a case by case basis in light of the relevant policies. Unfortunately, farmers who are planning changes and are concerned about nuisance actions have no guidance from the statute or the courts.

Likewise, the act does not provide protection in situations where an agricultural operation pollutes or causes a change in the condition of any stream or causes overflow onto neighboring lands. This Virginia Code provision applies regardless of the length of time the operation has been in existence. If the statute provided an exception for agricultural operations that are in compliance with state, local and federal water pollution regulations, the protection against nuisance claims would not be negated. Either the state did not consider the effect of this provision on farmland preservation or the state’s interest in preventing water pollution takes priority over its interest in such preservation.

Unlike some right-to-farm statutes, the Virginia act does not mention distribution of the burden of proof. In traditional nuisance actions, if the plaintiff shows an unreasonable interference, then the defendant has the burden of showing that he uses his land in a reasonable manner and does not adversely affect the plaintiff. A few states have included a presumption that farm operations are not nuisances if they meet the statutory requirements of the right-to-farm acts. Because the Virginia statute has no such provision, the defendant farmer probably has the burden of showing that the right to farm act bars the nuisance claim. However, following traditional evidence rules, the plaintiff probably has the burden of showing that the farmer’s activities are not afforded protection due to negligence, improper use or significant change in operation.

The Act prohibits a locality from passing an ordinance which declares agricultural operations existing for one year a nuisance on account of a change in condition in the locality. However, the prohibition may not affect zoning ordinances. Because localities are free to designate areas according to the uses they desire in an area, residents could pressure the governing body to zone farm operations protected from nuisance actions out of business.

---

94 VA. CODE ANN. Sec. 3.1-22.29(C) (1983).

95 Grossman & Fisher, supra note 80, at 133.

96 Id. Arizona, Oklahoma, Vermont and Washington have such a presumption.

97 Id.

98 Id. at 134. Requiring the adversary of one who relies on a statutory exception to prove the exception is a rule of evidence.

99 VA. CODE ANN. Sec. 3.1-22.29(D) (1983).

100 Grossman & Fisher, supra note 80, at 161.
farmland would combine with other pressures of urbanization to ensure conversion of such land.

PRESERVATION OF PRIME AGRICULTURAL LAND ACT

The preservation of prime agricultural land act establishes a state policy of encouraging the preservation of important farmlands in Virginia and requires certain state agencies to prepare plans to implement this policy. Each agency's plans must contain an impact statement detailing the effect the agency has on conversion of important farmlands. The Act establishes a committee on the preservation of important farmlands to review the effect of an agency's plans. The committee operates as a subcommittee to the Virginia Council on the Environment and annually reports its findings to the Governor and state legislature.

Government land use has a significant impact on the conversion of farmland. The state controls the location of highways, reservoirs and various public facilities which require large tracts of land. The adoption of a farmland preservation program by the state provides an effective tool for preventing the conversion of farmlands. However, state agencies must rigorously adhere to the farmland preservation policy in order to save farmland from unnecessary destruction. Because the Act merely adopts a policy of encouraging preservation, agencies could still ignore the conclusions of an impact statement and convert important farmland. Thus, the degree of protection that this statute affords farmland may depend on how committed a certain agency or an administration is to preservation.

The Act defines important farmlands according to production and soil classification. In general, important farmlands are those which have been or are producing and have soil types suited for agricultural use. Specifically, the Act defines three categories of farmland to which the preservation policy applies: prime farmland, unique farmland and farmland of statewide or local importance. The statute adopts the Soil Conservation Service definitions for prime and unique farmlands. Prime farmlands are those lands with little soil erosion which can produce crops with minimum use of fertilizers, pesticides and labor. Unique farmlands are lands other than

---

102 Id. at Sec. 3.1-18.6.
103 Id. at Sec. 3.1-18.7 (1983 & Supp. 1988).
105 VA. CODE ANN. Sec. 3.1-18.5 (1983).
106 7 C.F.R. Sec. 657.5(a), (b) (1988).
prime farmland which produce high-value food and fiber crops such as grapes, nuts, olives and fruits.\textsuperscript{107}

The statute requires localities to designate the important farmlands in their jurisdictions, making adequate provisions for nonagricultural uses.\textsuperscript{108} Apparently, a rural county could not designate all of its land as important farmlands in order to avoid future nonagricultural development. In addition, because important farmland does not include land committed to urban development, localities cannot designate a parcel of land as important farmland and stop its conversion once the development begins. However, deciding when a commitment to develop land has occurred is difficult. While the signing of a contract would more than likely satisfy the requirement, any lesser action may or may not constitute a commitment.

**AGRICULTURAL ZONING**

Agricultural zoning is a common method of land use control.\textsuperscript{109} Many states permit localities to adopt zoning ordinances that designate an area for either exclusive or nonexclusive agricultural use. Nonexclusive zoning allows agricultural and other uses within an agricultural zone; exclusive zoning permits only agricultural uses.\textsuperscript{110} Because nonexclusive zoning allows nonagricultural uses to exist and expand within areas labeled agricultural, its use is less likely to protect farmland than exclusive zoning.\textsuperscript{111} Even though exclusive zoning is a more effective preservation method, its validity is more difficult to sustain in court. In states that permit exclusive zoning, such as California and Wisconsin, courts have generally upheld the use of such ordinances in areas where little developmental pressure exists. However, development surrounding an exclusive agricultural zone could increase to the point where restricting land use in that area becomes unreasonable and thus invalid.\textsuperscript{112}

The Virginia zoning statute authorizes localities to classify lands within their jurisdiction according to various uses and to designate zones for the purpose of preserving agricultural and forest lands.\textsuperscript{113} The statute establishes a nonexclusive zoning system so that localities can permit various other uses in an area which is

\textsuperscript{107} VA. CODE ANN. Sec. 3.1-18.5.

\textsuperscript{108} Id. at Sec. 3.1-18.5(3).

\textsuperscript{109} Rose, Farmland Preservation Policy and Programs, 24 NAT. RESOURCES J. 591, 600 (1984).

\textsuperscript{110} Id.

\textsuperscript{111} Id. at 601.

\textsuperscript{112} Id.

designated as agricultural. Likewise, a locality can grant a special exception or use permit for nonagricultural uses in a pre-existing agricultural zone. Localities can grant these permits at their discretion because the statute does not require them to follow any standards.

The state gives localities a wide discretion regarding the zoning of land within their jurisdictions. The statute limits localities to designating areas only for certain purposes, but it makes no special provision which either encourages or requires localities to protect important farmlands from conversion. Thus, any preservation efforts based on zoning must come from the locality. However, localities cater to local concerns; they cannot effectively tackle a problem, like the conversion of important farmlands, which is statewide in scope. Indeed, the state is a better vehicle for preserving farmland because it has the jurisdictional authority to design a program for lands which cross local political boundaries.

The current zoning system in Virginia is not an effective method for preventing farmland conversion. Localities permit nonagricultural uses to exist and expand within agricultural zones and individuals can readily receive special use permits allowing development in agricultural zones. Virginia could certainly benefit from a detailed agricultural zoning program.

CONCLUSIONS

While several Virginia laws address farmland preservation, they do not prevent the conversion of important farmlands to nonagricultural use. Most of the programs focus on easing the burdens of approaching development for farmers rather than preventing conversion of agricultural lands. Deferred taxation, agricultural districting, agricultural zoning, the right to farm legislation and the preservation of prime farmlands act all protect farmland to some degree. However, each act could be amended to provide stronger incentives for farmers to keep farming, alleviate more developmental pressures and protect important farmlands from conversion. Indeed, the current Virginia programs provide a good base on which to build a comprehensive farmland preservation scheme which would protect this valuable economic and environmental resource.

If Virginia is to have a policy of preserving farmlands from conversion through tax relief measures, then the legislature needs to enact a tax relief program which

---

114 Id. at Sec. 15.1-491 (1981 & Supp. 1988).
115 Id. at Sec. 15.1-491(c).
118 Geier, supra note 105, at 686.
119 VA. CODE ANN. Sec. 15.1-491(c).
realizes that policy. If the legislature opts to keep the current system of deferred tax assessment, then it needs to address the problem of tax shifting. Some states have addressed this problem by providing subvention payments to the localities using state monies.\textsuperscript{120} Certainly the inclusion of subvention payments in the Virginia tax relief program would alleviate the problem of tax shifting. However, because the overall tax relief structure of deferred taxation is fairly ineffective as a farmland preservation technique, a more effective solution needs to be adopted.

The other major type of tax relief, circuit breaker tax credits, offers an alternative solution to the problem of tax depletion. Two states, Michigan and Wisconsin, place the burden of lost taxes directly on the state by allowing farmers to receive a dollar-for-dollar state tax credit for the local property taxes they pay.\textsuperscript{121} Yet, the circuit breaker credits, like deferred taxation, do not prevent land from being converted to nonagricultural uses.

One type of differential tax assessment, the restrictive agreement, appears to be the most effective tax relief technique for promoting the preservation of farmland. It focuses on preventing changes in land use rather than merely easing the burdens of encroaching development. In order to receive use value assessment, farmers must enter into contracts to maintain their land in agricultural use for a certain number of years.\textsuperscript{122} This assures the postponement of development and the preservation of economically and environmentally important lands.

The adoption of restrictive agreements along with subvention payments would provide a strong preservation incentive that avoids the tax shifting problem inherent in the current program. However, because tax relief is neither a comprehensive nor effective protection against the conversion of important agricultural lands, additional measures are necessary to slow the trend of changing land use in the state.

The agricultural and forestal districting act provides a good foundation upon which to build a program of land preservation.\textsuperscript{123} Modifying the act in certain areas would increase the protection afforded the state’s agricultural, forest and open space lands. Lowering the acreage requirement to 100 acres for the core parcel would provide protection for more lands. Loosening the production requirement for the receipt of benefits and extending the act to open space lands would further the policy of preserving land as an environmental resource rather than just for its economic input. Within a district, a limit on the annexation of land by municipalities would slow the conversion of agricultural and forest land bordering on cities and towns.\textsuperscript{124} Including a provision requiring district landowners to adopt sound conservation practices would

\begin{itemize}
  \item \textsuperscript{120} Dunford, \textit{supra} note 31, at 694.
  \item \textsuperscript{121} NALS, \textit{supra} note 1, at 19.
  \item \textsuperscript{122} \textit{Id.} at 18.
  \item \textsuperscript{123} Meyers, \textit{supra} note 38, at 646.
  \item \textsuperscript{124} NALS, \textit{supra} note 1, at 20.
\end{itemize}
increase the land’s production capabilities and economic profitability so that owners would more likely keep the land in agricultural use. Localities should incorporate agricultural districts in a zoning program which permitted only agricultural and compatible uses in districts. Thus, the legislature could do a good deal to amend the districting act so that it would more likely preserve land and more effectively protect the state’s farm economy.

The right-to-farm statute eases some of the burdens placed on farmers by encroaching urbanization. It bars nuisance claims and voids nuisance ordinances in many situations, yet its protection against such actions is far from complete. The number of farms producing cows and hogs in Virginia has declined in the past five years. In counties, such as Amelia, hog farms have all but disappeared in the last decade. The hog industry is more stable in counties such as Smithfield and Surry, yet the farmers there have no nuisance protection for actions initiated by rural neighbors. Removing the significant change exception would contribute both to the conservation of farmland and protection of a vital economic sector by enabling farmers to expand or improve their operations without penalty. Defining "improper" would give farmers better guidance for planning purposes. Including a provision that creates a rebuttable presumption that operations are not nuisances if they meet federal, state and local law regulations would clarify the burden of proof distribution and provide an easily-determined guideline for farmers planning their operations. Extending the Act to cover suits brought by rural neighbors would also better protect the farmer raising livestock.

The designation of important farmlands by the state and localities pursuant to the preservation of prime farmland act provides the basis for a farmland preservation program that affects conversion by private owners. If the state can use these guidelines to determine which lands it should avoid disturbing, it can use them to prevent conversion of important farmlands in general. The land classification system could be used in a state-mandated zoning program, a land banking system, acquisition of negative easements or other farmland preservation programs aimed specifically at land use control. Knowing which lands constitute the most important farmlands allows the state and localities to regulate development more efficiently. A land use control program guided by these designations could prevent important farmlands from being developed while allowing development of lands which are less suitable for agricultural uses.

Because the mechanisms of zoning are present in most localities, the state could easily implement such a program using the existing infrastructure. The adoption of a state-mandated agricultural zoning scheme would further Virginia’s policy of encouraging, promoting and preserving agriculture. The legislature should enact a measure which requires localities to place lands classified as important or prime farmlands in special agricultural zones. Development in those zones would be restricted to more intensive farming or farm support operations, with allowances for farm residences.

125 Statistics provided by the Va. Agric. Statistics Service. The number of farms with hogs decreased from 14,000 in 1984 to 8,500 in 1987. The number of farms with cattle and calves has decreased from 39,000 in 1984 to 38,000 in 1987.
The major countervailing consideration in establishing such a strict zoning program is that the more control the state has over land use, the less control individual owners have. Even if the courts would find zoning restrictions reasonable in light of the state policy of farmland protection, most landowners would strongly object to any such limitation on the use of their lands.

The creation of such a program would likely face stiff opposition from the legislature especially in areas where development pressure is high. So, a less restrictive alternative may be more feasible. The state could purchase development rights [PDR] for farmland in localities where rapid growth is occurring, thus preserving farmland and avoiding harsh restrictions on landowners. Hampton Roads, northern Virginia and the Richmond-Petersburg areas are ideal candidates for such a program.

Through either purchase or donation, the state could acquire negative easements (PDRs) on farmland which prevent the owners from developing those lands. The difference between the market value of land and its agricultural value would determine the cost of the development right for each farm. In areas with high development pressures, this would be very expensive. Spacing of easement purchases and limiting those to areas in danger of conversion could lessen the initial cost of preserving farmland in certain areas. The legislature could raise money for the purchase of development rights by bonding or levying taxes on either real property transfers, land grants or conversion of agricultural lands. A tax on property transfers would raise more money when development pressures are high, thus enabling the state to purchase more development rights when the need for farmland preservation is great. Likewise, a tax on conversion of agricultural land would both discourage conversion and raise revenues for purchasing development rights during times when they are most needed.

In addition to a PDR program, the state could exercise its power of pre-emption to save important lands already for sale from being developed. The state can substitute itself for the bona fide purchaser of important farmlands, acquire that land, place use restrictions on it and then either resell or lease it, giving preference to the bona fide purchaser. This would certainly be helpful during the initial stages of a PDR program and in cases where the owner of important farmland refuses to sell the property's development rights to the state.

Two other land use regulation techniques exist that Virginia could implement in order to expand its farmland preservation program: land banking and transfer of development rights [TDR]. In a land banking scheme, the state would acquire farmland

---

126 NALS, supra note 1, at 148.
127 Id. at 170.
128 Rose, supra note 110, at 617.
129 Id. at 617-18.
130 Id.
before it is for sale, place restrictions on its use and either resell or lease it. Such a program is well-suited for counties on the outskirts of the high development areas. Land prices in those areas have not increased dramatically, but the spread of urban growth is likely in the future. While land banking would further the preservation of farmland in Virginia and benefit the state through the increase in land values, its costs would be high. The state would have to establish an agency to handle the land bank programs and raise considerable amounts of money to purchase important farmlands. Even though the initial costs could be offset by later increases in the value of land and by benefits to the state's agricultural industry, the same preservation results could be achieved by a less-costly PDR program.

Under a transfer of development rights plan, development rights are purchased and then used in a different location. Virginia would have to divide its lands into development and preservation districts. Landowners in preservation districts have development rights for their lands, but they cannot develop their own lands. Instead, they can only sell the development rights to landowners in development districts so that those landowners can develop their lands at higher densities that zoning ordinances would permit. While this may be less costly for the state than a PDR program, the price would be paid by those people living in or near the development zones who could easily face overcrowding and congestion which would not have existed but for the TDR program. A TDR program is certainly a radical means of easing the cost burden on the state of obtaining development rights for important farmlands. Because of its novelty as a legal technique for farmland preservation, the legislature may be suspicious of such a program. Because of its side effects of overcrowding and congestion, landowners in would-be development zones may be greatly opposed to the measure. Thus, a TDR plan may be less attractive than a PDR program for Virginia.

The Virginia legislature needs to strengthen existing preservation programs and to adopt additional measures to protect the state's important farmland. The state should revise its tax relief program by entering into restrictive agreements with farmers and providing subvention payments to localities in order to reduce the financial impact of differential taxation. It should modify the agricultural and forestal districts act to include open space lands, integrate conservation practices and restrict annexation. The legislature needs to reword the right-to-farm statute to give it more substance. It also needs to use the land classification system established under the preservation of prime agricultural land act to halt farmland conversion caused by non-governmental actors. Incorporating the classification system in either a state-mandated zoning scheme or a

\footnotesize

131 Id. at 610.
132 Id. at 614.
133 NALS, supra note 1, at 174.
134 Id.
135 Rose, supra note 110, at 622.
PDR program would ensure the preservation of important farmland in the state. While a zoning scheme would be easier to implement given the existing zoning mechanisms throughout the state, a PDR program would not restrict farmers' use of their lands as much. Whichever plan the state legislature prefers, it needs to act now in order to slow the conversion rate of the important economic and environmental resource of farmland.
ENFORCEMENT OF THE SAFE DRINKING WATER ACT IN VIRGINIA

Mary A. Munson

INTRODUCTION

National enforcement of the Safe Drinking Water Act (SDWA) has been weak and ineffective, whether carried out by the Environmental Protection Agency (EPA) or a state agency. The situation may be due to inadequate funding for programs. Virginia reflects the national trend with its insubstantial record of enforcement. A citizen suit initiated by an environmental group has challenged the legality of the government’s enforcement efforts. If successful, the suit will overturn many administrative law cases that favor enforcement discretion in decision-making by agencies. However, the seriousness of the policy implications of non-enforcement may cause the courts to reexamine administrative law precedent.

Water pollution clean-up in the Commonwealth has focused on improving the quality of water sources, including groundwater and surface water sources such as rivers and lakes. These waters are not exclusively covered by the SDWA. The SDWA focuses on the improvement of public, rather than private, water supplies.

OVERVIEW OF THE SDWA

The Safe Drinking Water Act (SDWA)\(^1\) was enacted in 1974 to ensure that tap water reaching the public is safe for consumption. The statute requires that public water systems (PWSs) provide water which either does not exceed maximum contaminant levels (MCLs) or conforms to treatment requirements set out in the national primary drinking water regulations (NPDWRs).\(^2\) It also requires a permit for underground injections which may endanger underground sources of water, thus preventing excessive MCLs.\(^3\)

SDWA provisions allow PWSs to delay compliance under certain circumstances through variances\(^4\) and exemptions.\(^5\) These provisions were made more restrictive in the 1986 SDWA Amendments, wherein Congress provided that any water supplier applying for exemptions show that it has applied the best available technology (BAT) in attempting to comply.\(^6\) PWSs self-monitor by testing for contaminant levels and

---

reporting the results to the enforcement agency. The monitoring requirements have been described by one environmental group as "the heart of the law" because they enable the identification and correction of public health risks before they blossom into crises. They are also important because of budgetary constraints at the enforcing agency, which are likely to occur and which make monitoring by the government impractical.

The PWSs are defined as systems that have at least fifteen service connections, or regularly serve at least twenty-five persons. This definition covers not only public water systems which serve residential communities and are easy to identify and regulate, but may also include water suppliers that serve only a hospital, rest area, or campsite because they serve at least 25 persons. The EPA has treated these latter water suppliers as "non-community" water systems. The users of the water from these systems are transient, raising difficulties in notifying users about contaminated water. However, the non-community water system owner can be identified and regulated using regular PWS programs.

SDWA ENFORCEMENT IN VIRGINIA

Except for Wyoming, Indiana, and the District of Columbia, all other states and territories have been granted the responsibility for their own public water supply supervision. States receive this grant of primary enforcement responsibility ("primacy") by, inter alia, adopting regulations no less stringent than the NPDWRs and promulgating enforcement plans and emergency provisions. If a state fails to ensure enforcement, EPA should notify the state that a PWS is not in compliance, request the state to provide information about the supplier within fifteen days, and consider a civil action against the non-complying supplier.

Virginia obtained primacy by enacting its own PWS statute, which vests responsibility for ensuring safe drinking water in the Virginia Department of Health. Within the Department of Health, the Division of Water Programs oversees the drinking water regulations. The Virginia statute requires permits, minimum water quality standards, and mandatory testing of water for bacteriological, chemical, radiological,

---

8 DEAN, supra note 2, at 11.
11 42 U.S.C. Sec. 300g-2 (1982 & Supp. IV 1986) provides the authority, whereas the requirements are set forth in 40 C.F.R 142, Subpart B.
and other contaminants, and dictates that free technical assistance be provided to water suppliers at their request. The statute is consistent with the SDWA and conforms to the primacy requirements.

The Virginia Department of Health is responsible for all issues relating to public health, but shares responsibility with the State Water Control Board (SWCB) relating to regulating water quality. However, the Department has sole responsibility for regulating drinking water quality. Unlike the Department, the SWCB deals primarily with water issues, with authority to regulate sewage discharge into waters, investigate fish kills, put conservation measures in place, among other duties.

One of its most important duties is to establish and enforce water quality standards for state waters, including all surface and underground water within the state. By doing so, it carries out its responsibility to enforce National Pollution Discharge Elimination System provisions of the Federal Water Pollution Control Act. The State Water Commission estimates that out of the average 28 billion gallons per day of tap water used in Virginia, 27 billion gallons originate in state rivers and streams, and the remaining amount comes from the state's groundwater or 75 reservoirs. While the Health Department regulates the public water suppliers, the SWCB regulates activity that affects the quality of water received by the public water suppliers.

This dual oversight of state waters could lead to confusion and overlap of responsibilities. In 1983, a Water Study Commission recommended that a body be used to overview and coordinate the activities of all state entities considering water-related

---

14 VA. CODE ANN. Sec. 32.1-170 (1985).
15 VA. CODE ANN. Sec. 32.1-171 (1985).
16 The EPA has proposed a rule that would modify the primacy requirements for state Public Water Supply Supervision (PWSS) programs. The rule would require states already granted primacy to adopt new or revised EPA regulations. 53 Fed. Reg. 29194 (1988). This was necessary to accommodate the 1986 SDWA amendments, which require EPA to promulgate NPDWRs for 83 new contaminants.
18 Functions of the Department of Health include administering Medicaid, regulating medical care, licensing hospitals, day care centers and migrant labor camps, compiling vital statistics, and regulating sewage disposal as well as water treatment and supply. Report of Jt. Subcommittee, supra note 17, at 5-7.
19 VA. CODE ANN. Sec. 62.1-44.15 (Supp. 1988).
20 VA. CODE ANN. Sec. 62.1-44.15(3a) (Supp. 1988). A Circuit Court held that its primary responsibility was to protect groundwater from contamination in all forms. Environmental Defense Fund, Inc. v. Lamphier, 714 F.2d 331 (4th Cir. 1983).
issues.\textsuperscript{23} A Legislative Joint Subcommittee recommended that the Division of Water Programs, in particular, work with other agencies with which they interface.\textsuperscript{24} In fact, Virginia has received a grant from the EPA to develop a clearinghouse for groundwater data and has created a task force to help coordinate agencies.\textsuperscript{25} If the two agencies are coordinated, it may not be necessary to reorganize state water-related programs.

The Division of Water Programs carries out its mandate through six offices around the state. Each office is divided into water and wastewater issues. PWSs work with the District Engineers to achieve compliance with SDWA provisions. Most cities and counties have local health departments, which are responsible for the water works compliance. The state officials work with local owners and officials to achieve compliance and compile records.

If a violation is found, the state follows a set procedure. Notice is sent to the violator, followed by a "Passion Report" which identifies the type of violation, and any follow-up action. The enforcement officer in Richmond receives it and keys it into a computer and sends it to EPA on a computer disk.

**EFFECTIVENESS OF ENFORCEMENT IN VIRGINIA**

Enforcement efforts in Virginia reflect the national trend by being underfunded and insubstantial. The EPA maintains computerized records of PWS violations and actions taken by states and the EPA against the violators, which are called the Federal Reporting Data System (FRDS) reports. An analysis of these reports by the National Wildlife Federation (NWF) revealed an enormous gap between the number of reported violations and the number of enforcement actions taken. In Fiscal Year (FY) 87, for example, NWF found that out of 101,588 violations, representing 36,763 water systems (both community and non-community), only 2544 enforcement actions were taken by states.\textsuperscript{26} When the EPA discovers that a system is in violation, it must inform the state and take action on its own if the state does not commence action within 30 days of notification.\textsuperscript{27} According to the NWF report, the EPA took only 50 such actions in FY 87, which amounts to five one-hundredths of one percent of the 99,044 cases in which the state governments failed to take action.\textsuperscript{28}

---


\textsuperscript{24} Report of Jt. Subcommittee, supra note 17, at 16.

\textsuperscript{25} Interview with Evans Massie, Compliance Officer, Office of Water Programs, Virginia Department of Health (Sept. 1988).

\textsuperscript{26} DEAN, supra note 2, at 14.

\textsuperscript{27} 42 U.S.C. Sec. 300g-3(a) (1982 & Supp. IV 1986).

\textsuperscript{28} DEAN, supra note 2, at 17. According to the report, the enforcement figures are probably overstated. In the FRDS reports, the EPA may take several distinct "actions" for a given violation: a notice of violation, proposed administrative order, and a final administrative order. Thus the EPA can "triple count" enforcement data. \textit{Id.}
FRDS reports of EPA Region III states reveal that in FY 1987, Virginia's 4,768 PWSs experienced 2,578 violations, for which the state took 65 enforcement actions.29 This 2.52% enforcement rate is about average for all of the states.

It is conceivable that the lack of enforcement actions listed in FRDS does not mean that PWS violations continue in Virginia. EPA records only the state actions which constitute "appropriate" actions, because SDWA requires EPA to take action if "appropriate enforcement" is not taken by a state.30 Therefore, if the Health Department informally worked with a PWS, or compliance was actually achieved, SDWA goals could be attained without taking "action" as recorded in FRDS.

Evidence indicates that this may not be the case. In 1988, EPA performed an audit of the performance of non-community water systems, those who do not use a commercial or community water supply, by selecting a sample of non-community water supplies and scrutinizing their actual SDWA compliance performances. According to the audit report, EPA found "numerous" monitoring violations that were not reported in FRDS.31 In FY 86, there were 79 unreported coliform bacteria monitoring violations discovered at 51 non-community water supplies which recorded no violations in FRDS. The same year, EPA found that 86% of the non-community water supplies in Virginia did not monitor for nitrates as required by law, despite the fact that FRDS listed no monitoring violations at all for this contaminant.32

The audit also reported that in FY 86, Virginia PWSs were lax in initiating check sampling once a maximum contaminant level (MCL) violation occurs.33 Check sampling, required every 24 hours when MCL violations occur, is necessary to assure users that the contamination has been expunged.

The audit revealed other enforcement problems in Virginia. Upon confirmation of an MCL violation, the Health Department is required to notify the public of the problem.34 EPA found "little evidence" that the Health Department issued notification for any of the 36 MCL violations it found in the sample.35 It also found that data reported from Virginia to EPA was erroneous, to the extent that some counties were unaware that violations should be reported to the state.36 The discrepancies between reported violations and the data sent to EPA is understandable in light of the

29 Dean, supra note 2, at B-2.
32 Id. at 39.
33 Id.
35 EPA, supra note 31, at 40.
36 Id.
difficulties in coordinating the 1531 community and 2550 non-community water works in the state. If much of the field work necessary is done by county and city officials unfamiliar with the technicalities of the state-EPA reporting requirements, it is possible that many attempts to achieve compliance remain on the local and informal level.

Enforcement is complicated by the fact that violations are common in some circumstances where enforcement officials are reluctant to initiate action. One state official revealed that they have difficulties with PWSs without definable owners. For example, after a coal company emplaces a water supply, it may move out of the area and deed the rights to the supply to a new landowner, who may be unaware of his or her responsibilities under the SDWA. 37 Another problem is that the testing required in the monitoring provisions can be expensive for a small supplier. The State Corporation Commission must approve some water works' rate increases, but rate increases for smaller suppliers may be difficult.

Whether the problems with enforcement pose a threat to the health of Virginians is difficult to assess. It appears that contamination of drinking water sources has become a serious problem in some areas of the state, 38 so failure by the PWS to adequately treat water may raise health dangers. For example, problems with trilomethane contamination at PWSs in Northern Virginia have occurred, as well as high bacterial levels in limestone regions, mainly in the western part of the state. 39 A Health Department survey of tap water in the south-central Piedmont region revealed that 72.5 percent of the water had "serious water quality problems," the most prevalent being high levels of fecal coliform. 40 This constituent comes from human and animal fecal matter and is associated with gastroenteric infections. 41

ACTIONS TO COMPEL ENFORCEMENT

Enforcement of the SDWA is weak, but there is considerable uncertainty as to whether this deficiency can be remedied through judicial action. In the following months, a federal court will face this question in a suit brought by an environmental group. On December 12, 1988, the National Wildlife Federation issued a notice of its intent to sue the EPA. The notice contained charges that the EPA has failed to enforce standards, review state programs effectively, assist PWSs in violation of the SDWA, 37 Interview with Evans Massie, Compliance Officer, Office of Water Programs, Virginia Department of Health (Sept. 1988).

38 See WEIGMANN & KROEHLER, THREATS TO VIRGINIA'S GROUNDWATER (1988).

39 Interview with Evans Massie, Compliance Officer, Office of Water Programs, Virginia Department of Health (Sept. 1988).

40 WEIGMANN & KROEHLER, supra note 38, at 35.

41 DEAN, supra note 2, at 26.
adopt adequate enforcement policies, and prepare annual reports required by the SDWA.42

The major issue in the case is whether the EPA has discretion to refrain from taking enforcement actions for the provisions identified in the suit. The evidence that no enforcement actions have been taken against many violators is strong, because much of it is EPA documentation.

The language of some of the provisions supports the contention that EPA enforcement is non-discretionary. The specific provision reads:

Whenever the Administrator finds during a period during which a State has primary enforcement responsibility...that any public water system...

he shall so notify the State and provide such advice and technical assistance...as appropriate...to bring the system into compliance. 43

The SDWA then specifies that where a state fails to take enforcement action within 30 days of being notified of a violation, the Administrator "shall issue an order...or...commence a civil action...."44 The use of the word "shall" raises questions about the mandatory nature of the clause, where canons of statutory construction, common law precedents, and legislative history reveal divergent interpretations.

There is considerable weight in support of the interpretation that would make EPA's decision to issue an order or commence a civil action non-discretionary. The critical first step in statutory interpretation is "that language must ordinarily be regarded as conclusive."45 The ordinary meaning of the word "shall" implies that the action is mandatory. Accepted practice in drafting legislation is to use "shall" if an "obligation to act is to be imposed."46 The plain meaning of the word "shall" points to the mandatory nature of EPA enforcement.

The fact that the 1986 Amendments changed the word "may" to "shall" in the clause quoted above also supports this contention, since it implies a legislative recognition of the distinction between the words. Statements in the legislative history show that some legislators interpreted the language as mandatory.47 The wording appears to be clear and unambiguous, but extrinsic evidence suggests that "shall" may indeed be treated as a discretionary order.

---

44 42 U.S.C. Sec. 300g-3(a)(1)(B) (Supp. IV 1986). Other provisions contain the "shall" language as well, and raise similar questions about the mandatory nature of the sentence.
46 SUTHERLAND, 1A STATUTORY CONSTRUCTION 689 (1985). See also STATSKY, LEGISLATIVE ANALYSIS AND DRAFTING 133 (West 1984).
One commentator analyzed similar enforcement language in the Federal Water Pollution Control Act (FWPCA) and found that an overwhelming number of courts that interpreted the "shall" language held that the enforcing agency retained its enforcement discretion. The primary case cited was Sierra Club v. Train, where the Fifth Circuit examined a FWPCA provision very similar to that of the SDWA. The court said that although upon superficial examination it appeared that "shall" imposed a non-discretionary duty to act, extrinsic aids showed that the meaning of Congress was inconclusive, and "shall" need not be considered mandatory. Because many courts regard "shall" as a discretionary, rather than mandatory, command, the meaning is ambiguous.

When ambiguity exists in a statute, the administrative interpretation, although not conclusive, is entitled to deference if consistent with the statutory purpose and linguistically reasonable. If the EPA asserts that it has interpreted the "shall" language to mean "may," it is questionable whether it can be considered "linguistically reasonable," but numerous precedents making that interpretation may make it so. Assuming that the statutory purpose of the SDWA is to achieve safe drinking water within budgetary or resource constraints, EPA may argue that enforcement discretion is appropriate. Indeed, a spokesman for EPA has implied that budget realities are partially responsible for the enforcement deficiencies.

Another factor supporting the discretionary enforcement of agencies is the doctrine of separation of powers. In a speech given when signing the SDWA Amendments into law, President Reagan stated:

"I believe that Congress cannot bind the Executive in advance and remove all prosecutorial discretion without infringing on the powers of the Executive. It is unrealistic to expect that the EPA will ever have the resources or the need to take formal action against each and every violation of the Act, without regard to how trivial the violation or unfair an enforcement action would be."

The EPA may also argue that such a statutory interpretation is consistent with the general acceptance of the fact that enforcement decisions are so committed to agency discretion as to preclude review. The Third Circuit has suggested three criteria:

---


50 557 F.2d 485 (5th Cir. 1977).

51 Id. at 489.


54 22 WEEKLY COMP. PRES. DOC. 832 (June 19, 1986).
for determining whether review is appropriate, the second of which is the consideration of "the product of political, economic, or managerial choices that are inherently not subject to judicial review." Enforcement decisions may be so clearly related to economic, management, and political choices that they might be the appropriate decision to which the phrase "committed to agency discretion" might apply.

The National Wildlife Federation case outcome could depend on the court’s view of discretionary agency enforcement. If the environmental group succeeds in having a judge compel EPA enforcement, it may be necessary for EPA to request more money from Congress for SDWA programs, or divert money from other programs to programs of enforcement. Attention will be on this case, since another group has recently filed a lawsuit to compel enforcement of the Resource Conservation and Recovery Act (RCRA).

CONCLUSION

Virginia's weak enforcement of the SDWA reflects a national trend. Where public water suppliers remain in violation due to state inaction, the law requires the federal government to step in and take action to ensure that water is safe. The federal government has not been responsive. A suit has been filed to require EPA to increase its enforcement presence. The success of the suit may depend upon the degree to which courts are willing to abandon a traditional view of discretion in agency enforcement. If the suit is successful, Virginia's water system may benefit due to the additional presence of more certain sanctions against water suppliers who provide contaminated water.

55 Local 2855 v. United States, 602 F.2d 574, 579 (3d Cir. 1979).

56 The Environmental Defense Fund is alleging that RCRA provisions should be interpreted as meaning that EPA had a non-discretionary duty to perform certain tasks by the dates Congress had set forth. 19 Env't Rep. (BNA) 2376 (1989).
STATE CONTROL OF ARCHAEOLOGICAL RESOURCES ON PRIVATE LAND
David E. Clark

INTRODUCTION

In Titusville, Florida, road builders in a new subdivision were digging out a swampy bog near a pond. They gave up when, after digging twenty feet down with a backhoe, they had not hit the bottom of the bog. In the process, however, they discovered two skeletons. After the police determined no foul play was involved, archaeologists from Florida State University were called in and determined the bog was an Indian burial site approximately 8000 years old. The greatest discovery was in the skeletons; the bog had preserved them so well that many still contained the brains of the deceased. They are the oldest examples of human cellular structure anywhere. Windover Farms, Inc., donated the site to the state, paid some of the excavating costs, and lent its name to the site. The dig was to be closed in January 1987 with half the area left for future digs.¹

In Plymouth, Massachusetts, a bulldozer operator hired by a developer to work on a planned housing development graded an area adjacent to the development. In the process, he also destroyed a village, 800 years old, described as one of the last known Indian villages that existed before the arrival of the Pilgrims in 1620. Valerie Talmedge, director of the Massachusetts Historical Commission, said the site would have provided information on Indian life before European culture. The site had been left unmarked due to fear of looting of artifacts.²

These examples illustrate the contrasting treatment of America's archaeological resources. Public awareness of environmental problems is becoming more prevalent today, as evidenced by public involvement in pollution prevention, wildlife preservation, and the preservation of scenic wonders. Subsumed within this concern for our environment is an interest in the preservation of our cultural and historical past, the "human environment." One way of examining our cultural heritage is through the study of archaeology. This paper will survey the processes in which states and local governments have (and could) protect our archaeological resources, with an emphasis on the protection of archaeological sites on private land.

ARCHAEOLOGY IN GENERAL: THE PROBLEM

Archaeology is the science by which remains of ancient man can be methodically and systematically studied to obtain as complete a picture as possible of ancient culture


and society, thereby allowing past ways of life to be reconstructed. Ivor Noel-Hume and other historical archaeologists would also include the "recent past." Archaeologists generally attempt to examine how man lived and adapted the environment to his own use.

Today an archaeologist at work on a site is looking for all remains, not just obvious evidence of human activity. Not only artifacts (like pottery, metals, and stone tools) and features (permanent objects situated in the soil such as floors, pits, walls, foundations, post holes) are gathered and recorded, but also nonartificial materials are gathered: seeds, animal bones, and soils. These artifacts and features help identify the community that produced them both chronologically and culturally.

Archaeology is an important concern, according to noted archaeologist Jaquetta Hawkes, because it gives a people a "sense of having roots." Archaeology also helps chronicle our cultural heritage, by filling in the blanks left by the written record of history. At some point, every nation looks into its past with pride, and how successfully archaeology can preserve things today "may have a very real influence on how this nation thinks of itself in the centuries ahead." As the International Council on Archaeological Heritage Management proclaimed in its charter, "The protection of the archaeological heritage is the moral obligation of all human beings."

Given a motive for preserving archaeological sites, there is an ever-growing urgency to preserve them. One estimate calculated that in California alone two archaeological sites per day were being destroyed by construction projects, natural erosion, incompetent excavation, or just plain vandalism. Today, one can readily see the changing landscape through development just by driving through the country. This development may damage the ecological part of the environment, but destruction of archaeological sites is permanent. As Hester A. Davis of the Arkansas Archaeological Survey said, "You can't grow a new Indian site." Once one perceives the need to preserve such sites for the information they contain, however, subsidiary problems of identifying sites and analyzing which sites to preserve crop up.

---

3 M. JOUKOWSKY, A COMPLETE MANUAL OF FIELD ARCHAEOLOGY 2 (1980).
5 JOUKOWSKY, supra note 3, at 177.
6 NOEL-HUME, supra note 4, at 8.
7 Id.
9 G. MCHARGUE & M. ROBERTS, A FIELD GUIDE TO CONSERVATION ARCHAEOLOGY IN NORTH AMERICA 22 (1977).
10 Is There a Future for the Past?, Archaeology (Oct. 1971) (quoted in L. BRENNAN, BEGINNER'S GUIDE TO ARCHAEOLOGY 14 (1973)).
In the United States, there are potentially innumerable sites of archaeological significance. Indians have lived in North America since approximately 10,000 B.C.\(^{11}\) and, given the nomadic tendencies of some tribes, the possibility for sites anywhere is incredible. Also, James Deetz considers the potential number of historical archaeological sites in the eastern United States alone "astronomical".\(^{12}\) The number is astronomical because historical sites represent the period of maximum population in the U.S. The ultimate dilemma is in deciding which sites to save. In the past, homes of famous people were preserved, spawning "individual" archaeology.

However, archaeology studies cultures, not individuals; it would be dangerous to extrapolate generalizations about a society from a site used by an elite sector of that society.\(^{15}\) Archaeologists would thus ideally like to save all sites, at least until excavation. The Office of Technical Assessment of the U.S. Congress even says excavation should be the last resort in archaeological research because new techniques in research are constantly being developed.\(^{14}\) The key for state governments is to balance the common interest of mankind in the information contained in archaeological sites and society's interest in putting the land to an economically viable use.

**STATE LANDS**

In general, state governments control archaeological sites located on state owned or state controlled property. The state either explicitly reserves the right to investigate and excavate archaeological sites\(^{15}\) or grants permits to qualified institutions and individuals to excavate the sites, while the state retains ownership of any artifacts found.\(^{16}\) A state archaeologist or state historical commission usually has power to grant permits.

State governments usually provide for state agencies to "cooperate" with the state agency charged with protection of archaeological sites.\(^{17}\) The degree of cooperation, though, varies from state to state. Alaska, for example, provides for one of the more strict state programs on public construction sites. The Department of Natural Resources may survey an affected area for historical or archaeological sites before "public construction or public improvement of any nature is undertaken by the state, or by a governmental agency of the state, or by a private person under contract with or licensed

\(^{11}\) G. MCHARGUE & M. ROBERTS, *supra* note 9, at 62.

\(^{12}\) J. DEETZ, IN SMALL THINGS FORGOTTEN 32 (1977).

\(^{13}\) *Id.* at 30-31.

\(^{14}\) OFFICE OF TECHNICAL ASSESSMENT, CONGRESS OF THE UNITED STATES, TECHNOLOGIES FOR PREHISTORIC AND HISTORIC PRESERVATION 70 (1988).

\(^{15}\) See, *e.g.*, VA. CODE ANN. Sec. 10.1-901 (Supp. 1988).

\(^{16}\) See, *e.g.*, *id.* at Sec. 10.1-903.

\(^{17}\) See, *e.g.*, ALASKA STAT. Sec. 41.35.070 (Supp. 1988).
by the state or governmental agency of the state."\textsuperscript{18} The department has the power to stop construction and investigate and excavate the site if it determines that such sites will be adversely affected by the public construction or improvement. Of course, "[a]ll investigations, recording and salvage work shall be performed as expeditiously as possible so that no state construction project will be unduly impaired, impeded, or delayed."\textsuperscript{19}

Some states provide for inspection of state land before state sale and the possibility of retention of historic sites or structures for preservation.\textsuperscript{20} State governments overall do an admirable job in preserving state controlled archaeological sites. They recognize that they are preserving these sites and artifacts for the common welfare of the general public and consequently, they take their job seriously by mandating penalties for permit violations and vandalism. Unfortunately, the states do not extend their expansive protection to archaeological sites located on privately owned land.

PRIVATE LAND

IN GENERAL

State control of archaeological sites on private land can be described as limited; the states respect the sanctity of private property. State statutes provide for control of such archaeological sites (and artifacts found) by the landowner.\textsuperscript{21} Occasionally, a state will declare its wishes that a property owner will excavate or preserve a site in conformity with proper archaeological standards and methods. Oklahoma provides an example:

"[i]n order to protect and preserve historical, archaeological, and scientific information, matters and objects and other archaeological remains, which may from time to time be found on privately owned lands within Oklahoma, the Legislature declares as a statement of purpose that archaeological excavations on privately owned lands should be discouraged except in accordance with and pursuant to the spirit and authority of this statute."\textsuperscript{22}

Most states thus merely encourage a property owner to be civic minded; there are few state controls over the use of privately owned archaeological sites. Owners may loot a site for artifacts if they so desire without regard for their informational value. The ultimate control still resides in the property owner. States possess the power to acquire land, through gift or purchase, for the purpose of preserving archaeological or historic

\textsuperscript{18} \textit{Id.}

\textsuperscript{19} \textit{Id.}

\textsuperscript{20} \textit{See, e.g., IND. CODE ANN. Sec. 14-3-3.4-5 (Burns 1987).}

\textsuperscript{21} \textit{See, e.g., VA. CODE ANN. Sec. 10.1-904 (Supp. 1988).}

\textsuperscript{22} \textit{OKLA. STAT. ANN. tit. 53, sec. 361 (1988).}
sites. Very few states, however, take the extra step by allowing property acquisition through the process of eminent domain.

Another state tool is the conservation easement, a nonpossessory interest in real property for the purpose of preserving the archaeological aspects of the property, whereby the landowner maintains the site in accordance with the easement. The designated state agency or historical commission may declare a site or structure a historic site or landmark, thus ensuring state protection through various regulations. But if an individual owns the site, the state may only declare the site a landmark/historic site with owner consent. Easements must also be purchased from the owner. Consequently, the most common state mechanisms for protecting such sites are subject to owner consent, either through selling an interest or consenting to state control of the lands' archaeological resources.

PERMITS

Excavations on private lands could be controlled by the state through a permit system. Permits are generally required for excavations on state property and some states require a permit for excavating on private lands designated as state landmarks. Only a few states, however, have taken the next logical step and provided for a permit requirement for excavating on any land in the state. Colorado provides one of the few such examples:

"The society shall issue or deny permits for the investigation, excavation, gathering, or removal from the natural state of any historical, prehistorical, and archaeological resources within the state and determine whether or not the applicants for such permits are duly qualified to conduct investigations for which the permit is requested."

Such a statute, depending on the permit qualifications, could ensure excavations that conform with proper archaeological standards, yet protect the owner's property rights in the artifacts and his development rights.

The idiosyncrasies of specific state statutes may nevertheless present interpretive problems. One such ambiguity is in the North Dakota Code, which requires a permit for excavating on any land in the state. However, the Code also states that nothing in the chapter shall limit a private land owner from excavating on his own land or

---

23 See, e.g., TENN. CODE ANN. Sec. 11-6-114 (1987).
24 See ALA. CODE Sec. 41-9-242 (1982); ALASKA STAT. Sec. 41.35.060 (1988); HAW. REV. STAT. Sec. 6E-3 (1983); TEX. REV. CIV. STAT. ANN. art. 6081c (Vernon 1970).
25 See, e.g., VA. CODE ANN. Sec. 10.1-1009 (1988).
26 See, e.g., id. at Sec. 10.1-905 (Supp. 1988).
27 See, e.g., FLA. STAT. ANN. Sec. 267.12 (West 1988).
28 COLO. REV. STAT. Sec. 24-80-406(1)(a) (1982).
29 N. DAKOTA CODE Sec. 55-03-02.
someone with the owner's written consent. An attempted reconciliation of these provisions argues that the Superintendent (the permit issuer) cannot refuse to issue a permit where the applicant is the land owner or someone with the land owner's written consent.

Beck's reasoning, however, is directly contradicted by the Turley decision in New Mexico. Turley was hired by land owners to excavate an archaeological site on their land, using a front-end loader. The applicable New Mexico statute stated that no person could excavate an archaeological site with mechanical equipment on private land without a permit. However, nothing in the section was deemed "to require such owner to obtain a permit for personal excavation on his own land." Turley did not have a permit, though he did have a contract with the owners. Turley argued, however, that the word "owner" in the statute also included the owner's agent.

The Court of Appeals of New Mexico took a common sense approach to this argument and ruled that this section did not include the word "agent," whereas other sections did. By process of exclusion, therefore, the Legislature did not intend to include the owner's agent within the section. The Court of Appeals fixed on the word "personal" in the statute and determined the word meant that the act is done in person without intervention of another. "This statutory requirement of "personal" excavation cannot be reconciled with the contention that "owner" includes 'agent'", and concluded the common law rule of agency was not applicable to this statute. The court also said that permitting a land owner to excavate while requiring his agent to acquire a permit was consistent with the legislative intent of the statute: to preserve and protect structures, sites, and objects of historical significance in New Mexico and to discourage archaeology on private lands except in accordance with the provisions and spirit of the Act.

The Supreme Court of New Mexico then rejected this analysis and overruled the Court of Appeals, holding that Turley, as the land owner's employee, was not required to get a permit. The court asserted it was construing Section 18-6-11 "according to its

30 Id.
33 N.M. STAT. ANN. Sec. 18-6-11 (1978).
34 Id. at Sec. 18-6-9(B) (1978).
35 Turley, 96 N.M. at 594, 633 P.2d at 703.
36 Id.
37 Id.
38 Id. at 594, 633 P.2d at 702-3.
39 Id.
plain meaning.\textsuperscript{40} The court applied the law of agency, stating that a person may do an activity through an agent that he may do personally, "unless public policy or some agreement requires personal performance."\textsuperscript{41} The statute must expressly or by implication prevent an agent of the owner from acting.

The Supreme Court of New Mexico's interpretation of the statute nullifies its policy. The New Mexico legislature stated that "the historical and cultural heritage of the state is one of the state's most valued and important assets" and laid out a program designed to preserve the state's cultural heritage.\textsuperscript{42} The legislature expressly declared its intent to discourage private land owners from excavating archaeological sites except in accordance with the Act, including the permit system.\textsuperscript{43} The permit system is designed to assure that archaeological remains are removed with adherence to proper archaeological procedures, yet to preserve some discretion with the land owner. By allowing a land owner to hire anyone to excavate without a permit, the information the state seeks to preserve may be lost.

A strong argument can be made that the Supreme Court of New Mexico chokes on its own words: a strong public policy of preserving the state's cultural heritage mandates personal performance. Turley reveals the problems of accommodating private land owners within a permit system. State legislatures should be careful to be explicit in their permit statutes, lest more Turleys and circumventions of state policy arise.

COMPREHENSIVE LAND USE ACTS

The protection of archaeological resources can be attained by including such protection in legislation designed to protect the environment in general from the effects of development. California has enacted an elaborate statutory framework designed to protect its environment. The California Environment Quality Act (CEQA) provides that public agencies should not approve projects\textsuperscript{44} as proposed if there are feasible alternatives or mitigation measures which would lessen the significant environmental effects of such projects.\textsuperscript{45} The Act is primarily designed to apply to "discretionary" projects proposed to be carried out or approved by public agencies, including

---

\textsuperscript{40} Id.

\textsuperscript{41} Id.

\textsuperscript{42} N.M. STAT. ANN. Sec. 18-6-2 (1978).

\textsuperscript{43} Id. at Sec. 18-6-10.

\textsuperscript{44} Including direct activity by a public agency, activities undertaken by a person supported by any form of assistance by a public agency, and activities involving issuance to a person of a lease, permit, license, certificate, or other entitlement for use by a public agency. CAL. PUB. RES. CODE Sec. 21065 (West 1986).

\textsuperscript{45} Id. at Sec. 21002.
amendment of zoning ordinances, issuance of zoning variances, issuance of conditional use permits, and approval of tentative subdivision maps.\textsuperscript{46}

One part of the process is the filing of an Environmental Impact Report, a document to be considered by every public agency prior to its approval or disapproval of a project. The report provides public agencies and the public detailed information about the effect a proposed project will likely have on the environment. Additionally, it lists ways in which the significant effects of a project may be minimized and indicates alternatives to such a project.\textsuperscript{47} The public agency having principal responsibility for carrying out or approving a project has the responsibility for determining whether an environmental impact report or negative declaration shall be required for any project.\textsuperscript{48}

The state has authority to require preservation of "unique" archaeological sites or mitigation of development effects on the site.\textsuperscript{49} Examples of preservation include avoidance of such sites, conservation easements, and capping or paving the sites with a layer of soil before building. Mitigation involves excavation of the parts of the site that would be destroyed or damaged by construction. Such excavation should normally be completed within 90 days after final approval of a project.

The statute is instructive in establishing guidelines by defining a site as a "unique archaeological resource" if it has a particular quality, such as being the oldest or best example of its type, is directly associated with a prehistoric or historic event or person, or has information needed to answer important scientific research questions and there is demonstrable public interest in the information. The protection provided may be limited, however, because the statute expires January 1, 1994. One can expect another amendment to be forthcoming, though, because the statute was originally set to expire in 1986, but was amended.

Section 21083.2 may have been designed to address \textit{Society for California Archaeology v. County of Butte}\textsuperscript{50} in order to specifically protect archaeological sites. Prior to the decision, "environmental" was defined as "objects of historical or aesthetic significance." The court of appeals ruled the definition included archaeological sites.\textsuperscript{51} In this case, the Court overturned approval of a subdivision because the board of supervisors did not comment specifically and respond to the archaeological information contained in the environmental impact report. The board must state why certain positions or objections were accepted or rejected.\textsuperscript{52} Although objections based on the

\textsuperscript{46} \textit{Id.} at Sec. 21080.

\textsuperscript{47} \textit{Id.} at Sec. 21061.

\textsuperscript{48} \textit{Id.}

\textsuperscript{49} \textit{Id.} at Sec. 21083.2.

\textsuperscript{50} 65 Cal. App. 3d 832, 135 Cal. Rptr. 679 (1977).

\textsuperscript{51} \textit{Id.} at 837, 135 Cal. Rptr. at 682.

\textsuperscript{52} \textit{Id.} at 839, 135 Cal. Rptr. at 683.
The adverse effects of the development on archaeological sites were presented, the board may not have felt an obligation to respond due to the lack of explicitness in the statute. Section 21083.2 removes this ambiguity by mandating consideration of the effects of a project on archaeological resources during the approval process. Some jurisdictions provide protection for designated natural areas, such as coastal areas, scenic rivers, and "areas of critical state concern." Subsumed within these environmental protection acts are protections for archaeological resources. Such comprehensive frameworks are useful in identifying archaeological sites and mitigating the effects of development projects. This type of legislation would serve the purpose of collection of information contained in archaeological sites by preserving a site outright or at least by excavating the site properly within recognized archaeological procedures.

The likelihood, however, that other states will follow the lead of California and enact environmental protection statutes for the entire state is unlikely. A state is much more likely to pass such legislation covering a discrete natural area whose uniqueness or importance to the state is widely recognized, such as Virginia's Chesapeake Bay. Restricted, rather than statewide, areas of monitoring will also solve administrative problems by making monitoring easier and thus save already inadequate state archaeological funding.

**HISTORIC DISTRICTS**

Whereas regional environmental/archaeological protection plans are more suited to rural or undeveloped areas, the regulation of archaeological resources located on private lands in more developed areas or in the urban context is best achieved through historic district legislation. Several states provide for local designation of such areas or specifically list the areas in the statute. While the state has to obtain owner consent before a single site is named a landmark, if the site is within a homogeneous area of historical or cultural significance, the entire area may be designated as a historic district and be regulated through ordinances, subject to various methods of owner registration.

---


56 *FLA. STAT. ANN. Sec. 380.05* (West 1988).

57 See, e.g., *VA. CODE ANN. Sec. 15.1-503.2* (Supp. 1988).
approval. For example, Connecticut provides that two-thirds of the owners have to vote affirmatively for a proposed district in order for it to become a historic district. 58

Currently, most historical district statutes use cultural, architectural, and historical significance as criteria. 59 One could possibly read archaeology into these criteria, but few states explicitly specify archaeological resources as criteria to be evaluated in the designation process. 60 Nevertheless, historic district statutes and ordinances can easily be amended to require review of archaeological resources in the designation process.

The basic procedure in designating a historic district remains generally the same throughout the states. The process begins with the "historic preservation committee" that investigates and recommends landmarks or districts to the appropriate zoning or planning agency for designation as historic districts. The governing body then may pass ordinances or statutes designating the area as a historic district. The preservation committee then usually controls the district; its members are primarily represented by experts in history, architecture, and possibly archaeology. Any new structures to be built or changes to the exteriors of buildings in the district have to be approved by the commission with what is called a certificate of appropriateness. Louisiana is unique in prohibiting excavations without a certificate of appropriateness if earthworks of historical or archaeological importance exist in the historic district. 61 When the commission denies a certificate of appropriateness for modification or construction, the property owner can then appeal the decision to the appropriate governing body over the commission, and then to a court of competent jurisdiction. 62 Under the historic district procedure, the property owner's interest in a viable economic use of his property is protected, while the public's interest in maintaining the historical identity of the area is preserved.

As of 1983, 800 to 1000 communities in the United States have created preservation committees at the local level, a jump from a mere 11 in 1957. 63 The phenomenal jump can be credited in some part to the Supreme Court's decision in Penn Central Transportation Co. v. New York City. 64 The facts in Penn Central involved the city

58 CONN. GEN. STAT. ANN. Sec. 7-147b (West 1988).

59 New Hampshire uses "cultural, social, economic, political and architectural history". N.H. REV. STAT. ANN. Sec. 31:89a (1988).

60 See, e.g., IDAHO CODE Sec. 67-4607 (1980); ILL. ANN. STAT. ch. 34, sec. 6503 (Smith-Hurd 1988); MICH. COMP. LAWS ANN. Sec. 399.202 (West 1988); S.D. CODIFIED LAWS ANN. Sec. 1-19B-34 (1985).

61 LA. REV. STAT. ANN. Sec. 25:737 (West 1988).

62 Id.


designation of Grand Central Terminal as a landmark, thus subjecting it to unique regulations, and whether such designation constituted a "taking" under the Fifth Amendment. The Court generally held that state landmark designation, and thus historic district designation, does not constitute a "taking" under the Fifth Amendment. The Court noted it had upheld land use regulations for the health, safety, morals, or general welfare that nevertheless destroyed or adversely affected recognized real property interests. But, "a use restriction on real property may constitute a 'taking' if not reasonably necessary to the effectuation of a substantial public purpose...or perhaps if it has an unduly harsh impact upon the owner's use of the property."

In Penn Central's case, the Court held the landmark designation statute was not discriminatory zoning, since it was part of a comprehensive plan to preserve historic or aesthetic sites in the city, and the designation was not arbitrary since the owner had a right to judicial review. The Court concluded the landmark designation was not a "taking" because the restrictions were substantially related to the promotion of the general welfare and still permitted reasonable beneficial use of the site.

Historic districts are arguably in a better position than landmarks as a result of Penn Central. Penn Central did not dispute that a showing of property value diminution would not establish a "taking" if the restriction had been imposed as a result of historic district legislation. The Court also noted that duties imposed by zoning and historic district legislation that apply throughout particular physical communities provide assurances against arbitrariness.

One can infer from Penn Central that in order for historic district designations to be constitutional, they can not be arbitrary and must provide for reasonable beneficial use by the owner. The cure for arbitrariness is specific criteria. The ordinance should also permit the owner to enjoy the beneficial use of the property, either economically or personally. Consequently, with ordinances covering

---

65 Id. at 125.
66 Id. at 127.
67 Id. at 132, 133.
68 Id. at 138.
69 Id. at 131. The Court cited Maher v. City of New Orleans, 516 F.2d 1051 (1975), which upheld the Vieux Carre ordinance, an architectural control ordinance applicable to New Orleans' French Quarter, as a permissive means to pursue the legitimate state goal of preserving the "tout ensemble" of the Quarter. The property owner did not show a "taking" because he did not prove the ordinance denied him a reasonable rate of return on his property.
70 Penn Central, 438 U.S. at 135.
71 Alaska provides that a historic district is a compact area of historical significance in which two or more structures important in state or national history are located. "Structures important in state or national history" is defined as property listed in the National Register of Historic Places, characteristic of the Russian-American period before 1867, early territorial period before 1930, or reflecting indigenous characteristics of Native culture in Alaska. ALASKA STAT. Sec. 29.55.020 (1986).
archaeological sites, preservation or excavation of the site should not be duly unreasonable so as to destroy the owner's business, for example. In any event, the right to judicial review of a historic district designation and of denial of a certificate of appropriateness mitigates any damage the regulation might entail. Judicial review can adequately reconcile a property owner's desire to utilize his property with the public's interest in preserving its cultural heritage.

THE ALEXANDRIA ARCHAEOLOGY PRESERVATION CODE

The city of Alexandria, Virginia, has proposed an ordinance system which would require archaeological survey work on any site slated for development. The Code was first proposed in 1984; subsequent research by the Alexandria Archaeology Commission revealed that only five jurisdictions in the U.S. require archaeological work prior to development (including Santa Fe, N.M. and Dade County, Fla.). A draft code, written by Assistant City Attorney Ignacio Pessoa, was then sent to the City Council in June 1987. A revised draft was completed in March 1988; the ordinance is now further being revised before public hearings and a vote by the City Council.\(^{72}\)

The Archaeology Preservation Code requires an archaeological clearance before grading or excavation can be undertaken on a site, depending on the area in which the site is located.\(^{73}\) On receipt of an archaeological clearance application, the city archaeologist conducts or accepts an independent archaeological assessment of the activity. The assessment may include a history, extent and location of ground disturbances, previous acts of ground disturbance, type and quantity of predicted archaeological resources on the site, and analysis of the public value of the site relating to its cultural heritage. The city archaeologist shall determine in the assessment the probability of adverse effects on archaeological resources by the activity, the benefit to be derived from preservation of the archaeological resources, and an archaeology management plan to be performed by the city or the applicant to secure the public benefits identified.\(^{74}\) The Code provides criteria for determining the significance of archaeological resources on a site.\(^{75}\)

If the city archaeologist determines the proposed activity would have no adverse effects on any archaeological resources on the site, he shall issue an excavation permit. But if the activity will have adverse effects on such resources, the city archaeologist and applicant are to decide on a management plan. Management plans may include monitoring of the excavation, surface reconnaissance, and full-scale excavations and in

\(^{72}\) P. Cressey, The Director's Chair, 6 ALEXANDRIA ARCHAEOLOGY VOLUNTEER NEWS 2 (April 1988).

\(^{73}\) Old Town Alexandria, the designated historic district, entails the most protection, then parcels with buildings over one hundred years old on them, and so on. ARCH. PRES. CODE Sec. 7-6-344 to 346.

\(^{74}\) Id. at Sec. 7-6-348.

\(^{75}\) Id. at Sec. 7-6-350.
place preservation; all resources found on private land, however, remain property of the land owner. Once a management plan is determined, a permit is issued, conditioned on fulfillment of the requirements of the management plan. The applicant further has right of appeal to the Archaeology Preservation Code Committee, the City Council, and then to the Circuit Court of the City of Alexandria. Alexandria has thus formulated an ordinance system designed to protect its archaeological resources, patterned on established historic district ordinances as applied to structures. Whether the Code can survive the influence of the real estate lobby and pass in its present form, only time will tell.

CONCLUSION

Archaeology has slowly been recognized as a part of our cultural heritage deemed worthy of protection in the public interest. States have wholeheartedly protected archaeological resources on state lands, but have been reticent about treading on the rights of private property. Whether states will expand the principle that a private land owner cannot use his property so as to harm the public interest (and thus protect archaeological sites on private land) depends on the public interest. Extensive protection of such sites on private land is premised on the assumption that the public actually cares.

Given the proliferation of historic district commissions, one can safely assume there is growing interest in the preservation of our cultural heritage. Comprehensive permit procedures, comprehensive land use statutes recognizing archaeological preservation, and archaeology ordinances in the historic district setting should be enacted to further the public welfare. The examples in the introduction are merely typical; the discovery of archaeological sites in construction sites happens every week. The preservation of man's cultural heritage requires more control over the actions of private property owners. To allow a land owner to destroy or excavate a site without any control or procedures is like taking words out of context - the essential meaning is lost forever.

76 Id. at Sec. 7-6-351.
77 Id. at Sec. 7-6-352.
78 Id. at Sec. 7-6-368 to 70.
INTRODUCTION

When state and local governments establish land use restrictions that hinge, at least in part, on the nature of expression made on the premises, the police power cases and First Amendment precedents are both implicated. Yet the standards in the two areas are vastly different. This conflict often arises in the context of a challenge to a zoning law that restricts adult entertainment establishments. Beginning with Euclid v. Ambler Realty, the Supreme Court has developed standards for reviewing exercises of the police power in the form of land use restrictions. The Court has shown great deference to the states in this area, defining broadly the goals which it will consider permissible for state action and viewing tenuous connections between purpose and action as sufficient to pass constitutional muster. In the area of free expression, the precedents are much less favorable to the government. Speech, including conduct that is equivalent to speech, is generally afforded substantial protection.

How is the Court to analyze a case to which both of these lines of cases apply? Should it view the problem as a possible infringement on free expression and apply a high level of scrutiny? Is it more appropriate to determine the level of review and the questions to be asked by making reference to the power being exercised and, thus, give deference to the state?

This paper will briefly examine the history and development of each of these two areas of the law. It will then examine those cases that have come before the Court that have involved both areas. Finally, there will be a discussion of possible flaws in the Court's treatment of these cases and recommendations for resolving the conflict between these areas of the law.

Suggested changes in the Court's analysis include adoption of standards from First Amendment law rather than only the generally deferential tests that are found in the police power cases, abandonment of the "secondary effects" test for content-neutrality and a more realistic approach to the question of the availability of alternative avenues of communication.
THE ZONING POWER: ITS HISTORY AND ITS LIMITS

A BRIEF HISTORY WITH SELECTED PRECEDENTS

The first comprehensive land use regulation was that of the City of New York in 1916. The genesis of the statute was a conflict between the owners of fashionable shops and the owners of factories in the garment district. The garment district, historically confined to the southern part of the city, had been expanding and encroaching on the shop areas. The shop owners lobbied for legislation to confine the factories.

On March 25, 1911, the Triangle Shirt Waist Company fire killed over 140 garment workers. An investigation revealed that the factory had been poorly constructed and that all means of escape from the building had been locked from the outside to prevent employees from avoiding their work. This event turned the political tide in favor of the shop owners. The state legislature quickly gave the city zoning power and the ordinance came soon thereafter.

The Supreme Court did not address the validity of the power to zone for ten years after the passage of the New York ordinance. When it did, it was not the metropolis of New York that was challenged, but the Village of Euclid. The Ambler Realty Company had, for fifteen years, bought and sold land in Euclid. When it sold a parcel for residential use, it inserted a restrictive covenant in the deed, so that the land could only be used for residential purposes. At the same time, Ambler was keeping an island of unrestricted land. Ambler would eventually have had a stranglehold on the market for land that was available for commercial use.

Unfortunately for Ambler, Euclid passed an ordinance which, *inter alia*, restricted the uses of the land in some parts of the village. Potential uses of land were placed in a hierarchy. Parks and single-family residences were among the highest uses. The lowest class of uses included such things as garbage incineration facilities and prisons. The statutory plan was a "cumulative" one, meaning that land could be put to a higher use than it was zoned for, but not a lower one. If an area was zoned for the

---


3 *Id.* at 884.

4 *Id*.

5 *Id*.


7 See generally S. KURTZ & H. HOVENKAMP, supra note 2.

8 Euclid, 272 U.S. at 381 (there was another class of uses excluded completely).
highest uses, most commercial activities were prohibited. Significant restrictions were placed on some of the tracts owned by Ambler.

Ambler brought suit, alleging that it had held the land for years for the purpose of industrial development, that the land was especially adapted to such uses, and that its market value would be reduced by sixty-five to seventy-five percent by the restrictions. The Supreme Court held, 6-3, that the ordinance was a valid exercise of the state's police power and was not a taking without due process of law. The extent of legitimate power to zone depended on the circumstances in which it was exercised. Unable to draw a precise boundary, the Court pointed to the common law of nuisance as a guide. Justice Sutherland, writing for the Court, noted, "[a] nuisance may be merely a right thing in the wrong place, like a pig in the parlor instead of the barnyard." Deference was to be accorded to the judgment of the legislature. In fact, before the Court would declare a land use ordinance unconstitutional, it would have to be shown that the ordinance was "clearly arbitrary and unreasonable, having no substantial relation to the public health, safety, morals, or general welfare."

The result in this case is somewhat surprising, coming as it did during the "liberty of contract" or "Lochner" era. The Court had struck down many statutes that regulated employment on the theory that individuals had a right to make their own contracts. Yet in Euclid and in Welch v. Swasey, a case involving a challenge to a restriction on the heights of buildings, the Court upheld significant restrictions on the right to contract for construction and for the full use of one's land.

The cases in the two areas (land use and employment) could be distinguished on the basis of greater externalities in the land use context. The Court may have believed that an employment contract did not have substantial effects on non-parties,

---

9 Id. at 380-81.
10 Id. at 383.
11 Id. at 384-84.
12 Id. at 397.
13 Id. at 387.
14 Id. at 388.
15 Id. at 395.
16 S. KURTZ & H. HOVENKAMP, supra note 2, at 882.
17 Euclid, 272 U.S. at 365.
19 S. KURTZ & H. HOVENKAMP, supra note 2, at 883. Theories of externalities had become well-developed by 1926. The Court's apparent appreciation of the difficulty of the "free rider" problem (apartments that took advantage of the lower density of surrounding areas) lends more credibility to the suggestion that these theories motivated the Court. However, it is more likely that the observations as to "free riders" were based on common sense and intuition, not a conceptual theory.
while decisions as to land use did. A more likely explanation of the greater deference shown by the Court in the cases involving restrictions on land use is the long history of governmental regulations affecting land use. There have been such regulations in England for over 600 years.\textsuperscript{20} After the \textit{Euclid} decision, many municipalities imitated the statute that was challenged in that case. Cumulative zoning was widely used.\textsuperscript{21} Planners felt that residential areas needed protection from encroachment by commercial uses. However, the reverse, that commercial uses might need protection from encroachment by residential areas, was not a widely held view.\textsuperscript{22} Thus, the ideal system was one that kept commercial uses from burdening residential areas with their unwanted presence, but did not keep an individual or family from building a home among the factories.

In 1928, the Court decided \textit{Nectow v. Cambridge},\textsuperscript{23} a case that was similar to \textit{Euclid} in several respects: the plaintiff had land that had been zoned residential and he wanted to sell it for commercial use.\textsuperscript{24} Here, in contrast to the result in \textit{Euclid}, the Court found that the plaintiff's Fourteenth Amendment rights had been violated.\textsuperscript{25} The plaintiff's land, a narrow strip, was surrounded by industrial facilities. As zoned, it was worthless.\textsuperscript{26} Unlike \textit{Euclid}, this was a challenge to the statute as applied, not a facial challenge.\textsuperscript{27} The Court found that the application of the statute in this instance was not integral to the general plan and bore no substantial relation to the public health, safety, morals, or general welfare.\textsuperscript{28}

The trend today is away from "cumulative" zoning and toward "exclusive" restrictions. Under an "exclusive" zoning statute there is no hierarchy of uses. For each area certain uses are permitted. All others are excluded.\textsuperscript{29} Planners now believe that some of the uses that were thought of as lower under the cumulative schemes need protection from uses that were thought of as higher; that, for instance, some commercial uses need protection from residential development.\textsuperscript{30}

\textsuperscript{20} D. MANDELKER, LAND USE LAW 1 (1988). "Public control of land use has a long history." According to Mandelker, the Romans had building site restrictions in the fourth century, B.C. \textit{Id}.

\textsuperscript{21} S. KURTZ & H. HOVENKAMP, supra note 2, at 887.

\textsuperscript{22} \textit{Id.} at 887.

\textsuperscript{23} 277 U.S. 183 (1928).

\textsuperscript{24} \textit{Id.} at 187.

\textsuperscript{25} \textit{Id.} at 188-89.

\textsuperscript{26} \textit{Id.} at 187.

\textsuperscript{27} \textit{Id.} at 185.

\textsuperscript{28} \textit{Id.} at 188-89.

\textsuperscript{29} S. KURTZ & H. HOVENKAMP, supra note 2, at 887.

\textsuperscript{30} \textit{Id.} at 887.
LAND USE RESTRICTIONS V. CONSTITUTIONAL LIBERTIES

In the 1970's, the Court dealt with challenges to statutes restricting those who could live together. The Village of Belle Terre passed an ordinance that allowed only one "family" to live in the same house. A "family" could include any number of related persons, but could not include more than two unrelated persons.\(^{31}\) In *Village of Belle Terre v. Boraas*, the Court heard a challenge to the ordinance on the grounds of equal protection and the rights of association, travel and privacy.\(^{32}\) The majority analyzed the case as an exercise of the police power.\(^{33}\) It specifically found that no fundamental right guaranteed by the Constitution was involved.\(^{34}\)

In *Moore v. City of East Cleveland*, the plaintiff attacked a statute that made it illegal for her to live with her grandson.\(^{35}\) The Court distinguished the facts before it in *Moore* from *Belle Terre* on the ground that the *Belle Terre* statute affected only unrelated persons.\(^{36}\) The Court explicitly refused to extend the "usual judicial deference" to the legislature in *Moore* because the case dealt with a regulation of the family. Finding that such an intrusion threatened the right to due process under the Fourteenth Amendment, the Court held that neither *Belle Terre* nor *Euclid* governed this case.\(^{37}\) Although the facts of *Belle Terre* are quite similar to those of *Moore*, the Court's analysis in *Moore* was much more searching and less deferential. The reason for this difference is that the Court found that a fundamental right was implicated in *Moore*, while it explicitly found that none was involved in *Belle Terre*.

SUMMARY OF THE CONSTITUTIONAL LIMITS ON ZONING

States can delegate to their subdivisions the power to regulate the uses of land as part of their police power.\(^{38}\) This power can only be used for certain purposes, but these purposes are defined very broadly -- the public health, safety, morals, or general welfare.\(^{39}\) Such a broad definition does not make fertile ground for facial challenges

---

32 Id. at 7.
33 Id. at 7-9.
34 Id. at 7.
36 Id. at 498.
37 Id. at 498-99.
39 Id. at 395.
to zoning laws, but the possibility remains that the Court will find that no relation exists between a particular application of a law and one of the permissible purposes. A statute is also vulnerable if it infringes on a fundamental right.

FREE EXPRESSION

INTRODUCTION

The First Amendment protects, inter alia, freedom of speech and freedom of the press. These freedoms, when considered together, shall be referred to as "freedom of expression." Where the protection of these guarantees is sought, the Court must first determine whether there has indeed been any expression. This becomes an issue where a litigant engages in nonverbal conduct which he alleges is expression. If the Court finds that there has been expression which has been restricted, it must then decide if that expression is obscene and thus deserving of no First Amendment protection. If the expression is not obscene it may still be offensive, in which instance it will receive some protection, but not full protection.

ARGUABLY EXPRESSIVE CONDUCT

David Paul O'Brien burned his Selective Service registration certificate (draft card) on the steps of the South Boston Courthouse. Destruction of the certificate was a violation of federal law. In United States v. O'Brien, the Court expressly rejected the view that all sorts of conduct could be labelled speech even if the actor intended for the conduct to express an idea. The Court viewed the conduct in this case as having both communicative and non-communicative elements. Where such elements were found in the same conduct, "a sufficiently important governmental interest in regulating the non-speech elements can justify incidental limitations on" the speech elements. For

40 Id.
41 See Nectow v. Cambridge, 277 U.S. 183 (1928).
42 See Moore v. City of East Cleveland, 431 U.S. 494, 498-99 (1976). But cf. Village of Belle Terre v. Boraas, 416 U.S. 1, 7-10 (1974) (scrutinizing claim that rights of association, travel, privacy and due process were violated by a zoning ordinance). The Court ultimately found that no fundamental right was violated.
43 U.S. CONST. amend. I
45 Id. at 370-71.
46 Id. at 376.
47 Id. at 376, 381-82.
48 Id. at 376.
a government regulation under the police power that affects freedom of expression to be permissible, it must be within the constitutional power of the government and further a substantial state interest which is unrelated to the suppression of free expression. Any incidental restriction on First Amendment rights must be no more than what is necessary to further that state interest.49

The Court, in upholding O'Brien's conviction, distinguished the case from Stromberg v. California.50 In Stromberg, the Court struck down a statute that made it illegal to express opposition to organized government by displaying a flag, badge, banner or device.51 Chief Justice Warren, writing for the Court in O'Brien, found that the Stromberg case was distinguishable because the statute in that case was aimed at suppressing communication and could not, therefore, be upheld as a regulation of noncommunicative conduct.52

A year after O'Brien, the Court invalidated the suspensions of high school students for wearing black arm bands as a protest to United States policy in Vietnam.53 The Court found that such symbolic action was "closely akin to 'pure speech'" and therefore entitled to "comprehensive protection."54 The Court acknowledged that the school environment has special characteristics,55 but as long as the expressive conduct was not disruptive, it was entitled to First Amendment protection.56 "Undifferentiated fear or apprehension of disturbance is not enough to overcome the right to freedom of expression. Any departure from absolute regimentation may cause trouble."57

Thus, arguably expressive conduct may either be "symbolic" or it may be a mix of communicative and non-communicative elements. In the former case, the Court will view the conduct as virtually "pure speech" and afford it commensurate constitutional protection.58 In the latter case, the Court will apply the test of O'Brien, determining if the restriction at issue is within the power of the government under the Constitution, whether it furthers a substantial governmental interest, whether that interest is unrelated to the suppression of free expression, and whether any incidental restriction

49 Id. at 377.
50 Id. at 382 (distinguishing Stromberg v. California, 283 U.S. 359 (1931)).
51 283 U.S. at 361, 369-70.
52 O'Brien, 391 U.S. at 382.
54 Id. at 505-06.
55 Id. at 506.
56 Id. at 508.
57 Id.
58 Tinker, 393 U.S. at 503.
on constitutionally protected expression is limited to that which is necessary to further that interest.\textsuperscript{59}

UNPROTECTED SPEECH

Not all expression is protected by the First Amendment. Some expression is considered of too little value to be protected.\textsuperscript{60} In \textit{Roth v. United States},\textsuperscript{61} the Court heard two cases involving First Amendment claims. In the first case, Roth published and sold books, photographs and magazines. He was convicted of mailing obscene materials in violation of the federal obscenity statute.\textsuperscript{62} Alberts, the appellant in the other case, was convicted of violating a California statute by keeping obscene books for the purpose of selling them and by writing and publishing an obscene advertisement of them.\textsuperscript{63} In affirming the convictions, the Court declared that "ideas having even the slightest redeeming social importance...have the full protection of the \[First Amendment\] unless excludable because they encroach upon...more important interests."\textsuperscript{64} However, the Court specifically excluded obscenity from the area of constitutional protection.\textsuperscript{65} This holding required the Court to define obscenity.

That task proved to be quite difficult for the Court. In \textit{Roth}, the Court offered a test: if to the "...average person, applying contemporary community standards, the dominant theme of the material taken as a whole appeal[ed] to the prurient interest" in sex, the material was obscene.\textsuperscript{66}

The members of the Court could not, however, reach any lasting agreement as to the standard. In the ten years following \textit{Roth}, the thirteen obscenity cases that came before the Court produced fifty-five separate opinions. Five different views developed.\textsuperscript{67} One was shared by three justices. Warren, Brennan and Fortas laid out their test in \textit{Memoirs v. Massachusetts}.\textsuperscript{68} In their view, for material to be considered obscene, the government had to show that the dominant theme of the material, taken as a whole, appealed to the prurient interest in sex; that the material was patently

\textsuperscript{59} \textit{O'Brien}, 391 U.S. at 377.
\textsuperscript{60} \textit{E.g.}, \textit{Chaplinsky v. New Hampshire}, 315 U.S. 568, 572 (1942).
\textsuperscript{61} 354 U.S. 476 (1957).
\textsuperscript{62} \textit{Id.} at 480.
\textsuperscript{63} \textit{Id.} at 481.
\textsuperscript{64} \textit{Id.} at 484.
\textsuperscript{65} \textit{Id.} at 485.
\textsuperscript{66} \textit{Id.} at 489.
\textsuperscript{68} 383 U.S. 413, 419 (1966).
offensive because it affronted contemporary community standards relating to description or representation of sexual matters; and that it was utterly without socially redeeming value. This was the test generally followed by the lower courts.

Redrup v. New York marked the beginning of the Court's practice of issuing per curiam reversals of convictions for the sale or exhibition of materials which five or more members of the Court found not to be obscene. From 1967 to 1973, thirty-one cases were disposed of in this fashion. This became known as the Redrup approach.

Miller v. California was decided in 1973. The Court's opinion laid out a new, three-part test. Material was obscene if: the average person, applying contemporary community standards, would find the work as a whole appealing to the prurient interest; the work depicted or described, in a patently offensive way, sexual conduct that was specifically defined by the applicable state law; and the work as a whole lacked serious literary, artistic, political or scientific value.

Two facets of the Miller test are especially noteworthy. First, it rejects the "utterly without socially redeeming value" test of Memoirs. The standard it uses, lack of serious literary, artistic, political or scientific value, makes it much easier to find material obscene. Second, it allows local standards to enter the determination -- there is no national definition of obscenity. The Miller test is now the standard for obscenity.

OFFENSIVE SPEECH

If expression is not found to be obscene under the Miller test, it is entitled to constitutional protection. However, the Supreme Court has recognized a category of expression which is not obscene, but which is, because of its offensive nature, of too little value to receive the full protection of the First Amendment. In Chaplinsky v. New Hampshire, the Court listed the types of expression for which a citizen could be punished. Among these types were the "lewd" and the "profane." The Court found that such utterances were not needed for the communication of ideas and were of "slight social value as a step to the truth."

---

69 Id. at 419-20.
70 G. STONE, L. SEIDMAN, C. SUNSTEIN & M. TUSHNET, supra note 66, at 1121.
71 386 U.S. 767 (1967).
72 G. STONE, L. SEIDMAN, C. SUNSTEIN & M. TUSHNET, supra note 66, at 1122-23.
74 Id. at 24-25.
75 Id. at 30.
In 1968, Paul Robert Cohen was arrested for wearing a jacket with the words "Fuck the Draft" written on it. He was convicted of disturbing the peace by offensive conduct. The Supreme Court overturned the conviction. The state clearly could not punish Cohen for the underlying content of his message. Further, the state, in its role as guardian of the public morality, could not be allowed to remove from the public vocabulary words it deemed offensive. Because "[o]ne man's vulgarity is another's lyric," "the Constitution leaves matters of taste and style largely to the individual." A truly captive audience might have been entitled to protection from Cohen's expression, but the "ability of government, consonant with the Constitution, to shut off discourse solely to protect others from hearing it is... dependent upon a showing that substantial privacy interests are being invaded in an ... intolerable manner." This was so even though "[w]e are often 'captives' outside... the home." The emotive function of words was often the most important element of the overall message, so that a finding that the same cognitive thought could be expressed without that particular word would not extinguish a First Amendment claim.

In *FCC v. Pacifica Foundation*, the Court heard a challenge to the power of the Federal Communications Commission to enforce a content-based regulation against a radio broadcast that was "indecent but not obscene." The broadcast was a satirical monologue by George Carlin entitled "Filthy Words." The FCC prevailed. Offensive expression was "not entitled to absolute constitutional protection under all circumstances." Context was an essential consideration. Broadcasts entered the home, and because the audience could tune in at any moment, prior warnings could not

---

78 Id. at 26.
79 Id. at 18.
80 Id. at 23-25.
81 Id. at 25.
82 See id. at 21-22.
83 Id. at 21.
84 Id.
85 Id. at 26.
87 Id.
88 Id. at 750-51.
89 Id. at 747-48 (opinion of three justices).
90 Id.
provide complete protection to an unwitting listener. Nor was it a complete remedy that the listener could turn off the radio if he heard indecent language.\textsuperscript{91}

If the Court determines that expression is not obscene under \textit{Miller}, but is offensive, lewd or profane, it will examine the context in which the expression is found. If the audience is "captive," the Court is more likely to find that regulation of the expression is permissible. If there are youths present in the audience, particularly if they are able to be exposed to the expression without parental consent, the Court will be more willing to uphold a regulation of the expression.

WHEN WORLDS COLLIDE: THE POLICE POWER MEETS FREE EXPRESSION

The Court has issued a few landmark decisions in cases involving conflict between the police power, in the form of land use restrictions, and First Amendment values as embodied in adult entertainment. The variation in the analysis used in these cases and the inconsistency in the outcomes reveal that the conflict between police power precedents and the jurisprudence of free expression is a perplexing one for the Court.

Jacksonville, Florida passed an ordinance which declared that it was a public nuisance for a drive-in theater to show a movie which displayed any of several specified parts of the human anatomy if the picture was visible from any public place.\textsuperscript{92} The Court closely examined the arguments offered by the city in defense of the ordinance in \textit{Erznoznik v. Jacksonville}. The Court focused on the nature of the speech being regulated and used language from First Amendment jurisprudence.\textsuperscript{93} The Court declared that the ordinance was unconstitutional.\textsuperscript{94} The ordinance could not be upheld as a time, place and manner regulation because it was not content-neutral; nor was there a sufficiently captive audience to justify state intervention.\textsuperscript{95} The ordinance swept too broadly to be upheld as an exercise of the state's police power to protect children. The scope of the ordinance was not restricted to nudity which would be obscene, even as to children.\textsuperscript{96} The scope was too narrow for Jacksonville to successfully defend the statute on the ground that it was necessary to prevent motorists from being distracted. The Court found "...no reason to think that a wide variety of other scenes in the customary screen diet...would be any less distracting to the passing motorist."\textsuperscript{97}

\textsuperscript{91} \textit{Id.} at 748-49.
\textsuperscript{92} \textit{Erznoznik v. Jacksonville}, 422 U.S. 205, 206-7 (1975).
\textsuperscript{93} \textit{See id.}
\textsuperscript{94} \textit{Id.} at 217-18.
\textsuperscript{95} \textit{Id.} at 209-12.
\textsuperscript{96} \textit{Id.} at 212-14.
\textsuperscript{97} \textit{Id.} at 214-15.
A deeply divided Court upheld a Detroit zoning ordinance against a First Amendment challenge in *Young v. American Mini-Theatres*. The challenged statute classified certain enterprises, including adult theaters, as "regulated uses." It required that an adult theater not be located within 1000 feet of any two other "regulated uses" or within 500 feet of a residential area. Justice Stevens, writing for four members of the Court, noted, "[f]ew of us would march our sons and daughters off to war to preserve the citizen's right to see 'Specified Sexual Activities' exhibited in the theaters of our choice."  

Although the state could not validly enact a total suppression of this type of communication, "[t]he State may legitimately use the content of these materials as the basis for placing them in a different classification from other motion pictures." The remaining question was whether this classification was justified by any interest of the city. Stating that "[t]he city's interest in attempting to preserve the quality of urban life is one that must be accorded high respect," and that the city should be allowed to "... experiment with solutions to admittedly serious problems," Justice Stevens found that the classification was justified. The plurality opinion admitted that the restriction was content-based, but found it constitutionally permissible because it was viewpoint neutral.

Justice Powell concurred and found that First Amendment concerns were implicated only incidentally. He saw no denial of a full opportunity for expression to convey its desired message nor of a full opportunity for everyone to receive it. The rest of Powell's opinion provided a look at things to come. He found that Detroit had made no effort to suppress free expression. The case, he argued, was not to be seen as involving a content-based restriction on the time, place or manner of

---

99 *Id.* at 52. *See also id.* at 53-54.
100 *Id.* at 70.
101 *Id.* at 70-71.
102 *Id.* at 71.
103 *Id.* at 71-72.
104 *Id.* at 70.
105 *Id.* at 73 (Powell, J., concurring).
106 *Id.* at 78-9 (Powell, J., concurring).
107 Compare *id.* at 73-84 (Powell, J., concurring) with *Renton v. Playtime Theaters*, 475 U.S. 41 (1986) (each giving deference to state actor and lending credence to claims of neutrality toward content).
108 *Id.* at 80-81 (Powell, J., concurring).
communication. Rather, the city had decided to treat certain movies differently because they had a different effect on their surroundings.\(^{109}\)

The dissenting opinion of four of the justices argued that if First Amendment protection applied only to expression which more than a few of us would go to war to defend, then popular opinion would define the scope of an essential part of the Bill of Rights.\(^{110}\) This was counter to the purpose of the Bill of Rights, which, according to the dissenters, was "designed to protect against precisely such majoritarian limits on individual liberty."\(^{111}\) The dissent also noted the significant similarities between the facts of this case and those of *Erznoznik*, decided only one year earlier.\(^{112}\)

The statute challenged in *Schad v. Borough of Mount Ephraim* excluded all live entertainment from the borough.\(^{113}\) The appellants operated an adult bookstore. The store contained coin-operated devices which showed adult films. When the operators of the store added a coin-operated device that allowed a customer to see a live performance by a nude dancer behind a glass panel, they were convicted of violating the prohibition on live entertainment.\(^{114}\)

The Court found the convictions unconstitutional.\(^{115}\) Entertainment, including nude dancing, was entitled to First Amendment protection.\(^{116}\) The right assertedly threatened, not the power being exercised, determined the standard of review.\(^{117}\) When a zoning law infringed on a First Amendment right, it had to be narrowly drawn and had to further a sufficiently substantial state interest.\(^{118}\)

\(^{109}\) *Id.* at 80-82 (Powell, J., concurring).

\(^{110}\) *Id.* at 86 (Stewart, J., dissenting).

\(^{111}\) *Id.*

\(^{112}\) *Id.* at 88 (Stewart, J., dissenting) ("The factual parallels between that case and this one are striking. There, as here, the ordinance did not forbid altogether the 'distasteful' expression but merely required an alteration in the physical setting of the forum. There, as here, the city's principal asserted interest was in minimizing 'undesirable' effects of speech having a particular content. And, most significantly, the particular content of the restricted speech at issue in *Erznoznik* precisely parallels the content restriction embodied in...Detroit's..." ordinance. "In short, *Erznoznik* is almost on 'all fours' with this case."). *Id.*

\(^{113}\) 452 U.S. 61, 63, 64 (1981).

\(^{114}\) *Id.* at 62-64.

\(^{115}\) *Id.* at 65, 77.

\(^{116}\) *Id.* at 65-66.

\(^{117}\) *Id.* at 68 (citing *Thomas v. Collins*, 323 U.S. 516, 529-530 (1945)).

\(^{118}\) *Id.* at 68.
The Borough claimed that the ordinance was part of its plan to create a commercial area that catered only to immediate needs of residents. However, many of the permitted uses went well beyond such a purpose. Under Mt. Ephraim's statutory scheme, "[v]irtually the only item or service that [could] not be sold in a commercial zone [was] entertainment...." Further, the Borough failed to show that problems associated with live entertainment were greater than those associated with permitted uses, so that the ordinance could not be upheld as a reasonable restriction on time, place or manner.

In addition, the ordinance did not leave open adequate alternative avenues of communication, a requirement for it to be upheld as a valid time, place or manner restriction. The Court suggested in dicta that if the county had been the zoning authority and it had excluded live entertainment from the Borough, but not from the entire county, the restriction might have been upheld. The Court also indicated that its ruling might have been different if the Borough had established that live entertainment was available in reasonably nearby areas.

The Court made its most recent pronouncement on the state of the law in this area when it handed down City of Renton v. Playtime Theatres, Inc. The ordinance was designed to concentrate on adult theaters by forbidding them to locate within 1000 feet of any residential zone, single- or multiple-family dwelling, church, park or school. Writing for the six-member majority (Blackmun concurred in the result), Rehnquist viewed the resolution of the case as being largely dictated by Young v. American Mini-Theatres. Because the ordinance did not ban adult theaters completely, the Court decided that it should be analyzed as a time, place or manner restriction. Finding that the ordinance was aimed only at "...the secondary effects of such theaters on the surrounding community," and "...not at the content of the films," the majority ruled that the regulation was content-neutral and, as such, was acceptable if it was

119 Id. at 72.
120 Id. at 73.
121 Id.
122 Id. at 73 ("[P]roblems that may be associated with live entertainment, such as parking, trash, police protection, and medical facilities.").
123 Id. at 75-77.
124 Id. at 76.
125 475 U.S. 41 (1986).
126 Id. at 52.
127 Id. at 43.
128 Id. at 46.
129 Id.
designed to serve a substantial government interest and did not unreasonably limit alternative avenues of communication.\textsuperscript{130}

The Court found that the ordinance was designed to serve a substantial state interest.\textsuperscript{131} The ordinance stated that its purpose was to "prevent crime, protect the city's retail trade, maintain property values, and generally 'protec[t] and preserv[e] the quality of [the city's] neighborhoods, commercial districts, and the quality of urban life'...".\textsuperscript{132} Rehnquist noted that the ordinance did not declare its purpose to be the suppression of free expression. He also opined that the city would have tried to close the theaters or restrict their number instead of just restricting their location if it had been trying to suppress their message.\textsuperscript{133} The majority opinion buttresses its assertion that the purposes of the ordinance were permissible by pointing to Renton's reliance on the experiences of Seattle and Detroit and on its own "findings" from public hearings.\textsuperscript{134}

Playtime Theatres contended that the ordinance was underinclusive because it did not regulate other adult businesses that were likely to produce secondary effects similar to those of adult theaters. To rebut this contention the majority pointed to the lack of evidence of any such businesses in Renton. The Court found no reason to assume that Renton would not amend its ordinance if the need arose.\textsuperscript{135}

The majority found that the ordinance easily met the constitutional requirement of providing for adequate alternative avenues of communication. The ordinance left 520 acres, about five of the area of the city, as potential sites for adult theaters. The majority was not persuaded by the appellants' arguments that much of the land was already occupied, that almost none of it was for sale or lease and that there were no commercially viable adult theater sites within the specified 520 acres. "That respondents must fend for themselves in the real estate market, on an equal footing with other prospective purchasers and lessees, does not give rise to a first amendment violation."\textsuperscript{136} The government does not have to ensure that a speech-related business will be able to get a site at a bargain.\textsuperscript{137}

\textsuperscript{130} Id. at 47-50.
\textsuperscript{131} Id. at 50-51.
\textsuperscript{132} Id. at 48.
\textsuperscript{133} Id.
\textsuperscript{134} Id. at 50-52.
\textsuperscript{135} Id. at 52-53.
\textsuperscript{136} Id. at 54.
\textsuperscript{137} Id.
SOURCE OF STANDARDS

The standards for review of government actions should be both those raised by the right implicated and those tied to the power exercised. Different powers of government pose different dangers and potentials for abuse. Thus, challenged exercises of those powers should be held to a minimum standard, regardless of the context in which they are exercised. In addition, different rights have different roles, different levels of importance and different vulnerabilities. In recognition of this, when an exercise of governmental power infringes on a constitutionally protected right, a standard of review commensurate with that right should be imposed.

To be more specific, the zoning power is broad. The purposes for which it may be employed are broad. However, if a statute, or a particular application of a statute, does not serve these purposes, it should be struck down. Where the exercise of zoning power infringes on a constitutional liberty, such as freedom of expression, a "permissible purpose" should be a necessary, but not sufficient, condition for upholding the state action.

Imposing only the standard attached to the power involved, without reference to the freedom infringed, would allow the government, by being selective about how it enforces its will, to select a low standard of review for its own actions, even where the most vital guarantees for personal liberty are involved.

To impose only the standard connected with the right implicated might allow the government to employ its powers in dangerous ways. It would also give government a free reign in areas which involve no freedom for which the Court acknowledges constitutional protection.

SUGGESTED ANALYSIS

POLICE POWER. When the Court encounters a case in which a zoning ordinance impacts an adult entertainment establishment, it should first determine whether the ordinance is a valid exercise of the police power. The Court should require the individual challenging the statute to show that the statute is "clearly arbitrary and unreasonable, having no substantial relation to the public health, safety, morals, or general welfare." If the person making the challenge can make this showing, then the

---

138 The proposed analysis begins from the premise that the precedents in each area are correct interpretations of the Constitution.

139 The proposed analysis is suitable for any case in which an exercise of the police power is alleged to infringe on free expression. Challenges to restrictions on adult entertainment establishments are simply a common context for such to arise.

case should be over -- the statute should be declared unconstitutional. If the person making the challenge can make this showing as to a particular application of the statute then the Court should enjoin that application. 141

DOES THE ORDINANCE RESTRICT "SPEECH"? If the statute survives the "police power" test above, then the Court should inquire whether the conduct affected by the statute is, in fact, speech. If the conduct is verbal, the Court should generally have no difficulty finding that it is speech. If the conduct is nonverbal, the Court may find either that it is not speech at all, that it has both communicative and noncommunicative elements, 142 or that it is symbolic -- nearly "pure speech." 143 If it is not speech, the statute should be upheld.

If the conduct has both communicative and noncommunicative elements, the Court should apply the three-point test of O'Brien: whether the statute is within the constitutional power of the government; whether the purpose of the statute is unrelated to the suppression of free expression; and whether any incidental restriction on First Amendment rights is no more than is necessary to further that governmental interest. 144 If the statute passes all three prongs of this test, the Court should uphold it. If it fails any of them, or if the conduct is symbolic, then the Court should progress to the next step in the analysis.

OBSCENITY. If the case has not been resolved by the foregoing analysis, the Court should determine whether the speech involved is obscene. It should make this determination by applying the test put forth in Miller v. California: whether the average person, applying contemporary community standards, would find that the work as a whole appealed to the prurient interest; whether the work is a patently offensive depiction of sexual conduct that has been specifically defined by the relevant state law; and whether the work as a whole lacks serious literary, artistic, political or scientific value. 145 If the communicative conduct that is restricted meets all of these criteria, it is unprotected. 146 The person challenging the statute has no more claim than he or she would if there was no speech. The statute should be upheld.

OFFENSIVE SPEECH. If the Court finds that the expression is not obscene, it should decide whether it is profane, libelous, or insulting. 147 If it is not, the Court should

141 Nectow v. Cambridge, 277 U.S. 183 (1928).
144 O'Brien, 391 U.S. at 377.
146 Id. at 23.
proceed to the next step which is in the next paragraph. If the expression is offensive, lewd or profane, the Court should examine the context of the expression.\(^{148}\) The factors that the Court examines would include the presence of youths in the audience, the existence of a "captive" or unwilling audience and the expectations of those who are exposed to the expression unwittingly.

In the usual adult entertainment setting there will presumably be no youths exposed to the expression. No one will have the expression thrust upon them, and it is unlikely that anyone entering the establishment will be surprised by the nature of what they observe. Such a set of facts do not make a case for regulation.

CONTENT-NEUTRALITY. If the case has still not been resolved, the Court should determine whether the statute can be upheld as a valid restriction on time, place or manner. The Court should first ask whether the statute is content-neutral. Perhaps the most glaring error in the Court's analysis in \textit{Renton} was its use of legislative cognizance of the "secondary effects" of a type of speech as a basis for finding that the regulation enacted was content-neutral. The ordinance in \textit{Renton} was content-based on its face. Only "adult" theaters were restricted. A theater was only an "adult" theater if the films it showed were "distinguished or characterized by an emphasis on matter depicting, describing or relating to 'specified sexual activities' or 'specified anatomical areas'...." The lawmakers' concern for secondary effects is relevant -- it goes to the magnitude of the government interest involved and possibly to how narrowly tailored the law is as well. But it has nothing to do with classifying what type of ordinance it is that has been passed. Judicial use of legislative concern for secondary effects of communication with a certain content should be limited accordingly. If the statute is not content-neutral, it is not a valid time, place or manner restriction and should be struck down.

GOVERNMENTAL INTEREST. If the statute is content-neutral, the Court should decide whether it is designed to serve a governmental interest that is significant. Secondary effects are relevant to this determination.\(^{149}\) Of course, it would be naive to suppose that the city would state an impermissible purpose in the statute itself. Yet

---


\(^{149}\) Many of the "secondary effects" that the Court has shown concern for are actions of those who receive the communication. There are limited instances in which the tendency of a communication to cause those who receive it to take certain actions can serve as the basis for restricting that communication. The "fighting words" doctrine of \textit{Chaplinsky} is an obvious example. A more relevant example is Kingsley Int'l Pictures Corp. v. Regents of University of New York, 360 U.S. 684 (1959), striking down a statute that prohibited the issuance of a license to show non-obscene movies that "portray acts of sexual immorality [as] desirable, acceptable, or proper patterns of behavior". \textit{Id.} at 687. "The state, quite simply, has struck at the very heart of constitutionally protected liberty." \textit{Id.} at 688. "Advocacy of conduct proscribed by law is not... a justification for denying free speech where the advocacy falls short of incitement and there is nothing to indicate that the advocacy would be immediately acted on." \textit{Id.} at 689. Most adult entertainment is at least arguably not advocacy at all, but this should only serve to weaken the argument for making the communicator subject to restrictions based upon the actions of the recipient of the communication.
this is precisely what the Court did in Renton. The Court's credulity should certainly have been overcome by the fact that the provision explaining the purpose of the Renton ordinance was added to the ordinance after the lawsuit was instituted. It is also naive to believe, as the Court claimed to, that city lawmakers are so unsophisticated that they will not try to pursue impermissible purposes through indirect means where the direct means are plainly more likely to be successfully challenged. Such deference is inappropriate in this analysis. Further, before giving weight to post hoc claims of reliance on the experiences of other communities, the Court should weigh the relevance of those experiences to the situation at hand, the similarity of the measure at issue to the measures adopted by the communities whose experiences are allegedly relied on and the evidence that those who claim reliance actually knew of the experiences.

NARROWLY TAILORED AND ADEQUATE ALTERNATIVES. If a significant government interest is served, the Court must decide whether the statute is narrowly tailored to serve that interest. If it is not, the statute is unconstitutional.

If the statute is narrowly tailored to serve a significant state interest the Court must ask whether it leaves open adequate alternatives for communication. Theoretical availability of alternative avenues should be deemed insufficient. Even practical alternatives should not always be enough. If, for instance, an adult theater may open, but it must pay a great deal more for rent because of government restrictions, or it can get an economically feasible site, but the site is inferior (in terms of location, size and suitability for the type of structure needed) to what would be available absent content-based regulations, there must come a point where the alternative is not "adequate."150 "'One is not to have the exercise of his liberty of expression in appropriate places abridged on the plea that it may be exercised in some other place.'"151

The Court's treatment of this issue in Renton was sorely lacking. The Court assumed away the issue by stating "that [adult entertainment establishments] must fend for themselves in the real estate market, on an equal footing with other prospective buyers and lessees, does not give rise to a First Amendment violation"152 assumed away the issue. The operators of adult theaters are not on an equal footing with other prospective land users when they are subject to ordinances that restrict the possible

---

150 See generally Note, Zoning and the First Amendment Rights of Adult Entertainment, 22 VAL. U. L. REV. 695, 717-23 (1988). The author suggests that after determining that there is some land which an adult entertainment establishment could operate on within the ordinance, a court should determine whether the sites are commercially viable. If the zoning ordinance unreasonably restricts access to commercially viable sites, it should be struck down. If the ordinance does not have an unreasonable effect on access to commercial viability, a court should determine whether land is economically available -- whether it can be purchased or leased for a reasonable price. Although the author of the note refuses to say that any of these factors -- geographic availability, commercial viability or economic availability -- is dispositive, he asserts that they should all be considered.

151 Young v. American Mini-Theatres, 427 U.S. 50, 86 n.6 (1976) (Stewart, J., dissenting, quoting Schneider v. State, 308 U.S. 147, 163 (1939)).

152 Renton, 475 U.S. at 54.
locations for their businesses and which do not apply to the same extent to other possible land users.
INTRODUCTION

The former common law prohibitions against allowing recovery for negligently inflicted emotional harm have been eroded in some jurisdictions by abandoning arbitrary rules of law. In part, this development recognizes that the reasons for not allowing recovery were possibly overstated or imaginary. The present more liberal situation carries the dangers of compensating some questionable claims and overextending defendant's liabilities. Some commentators have suggested new non-arbitrary rules of law, aimed at correcting potential abuses while allowing recovery to deserving plaintiffs. The problem is to design a system which selectively separates meritorious from specious claims, emphasizing the role of judges.

Most of the published commentary involves two kinds of cases: 1) Emotional harm to a bystander who observes a victim's physical injury caused by a defendant's negligence; or 2) emotional harm to the plaintiff himself caused by a defendant's negligence but without any contemporaneous physical injury either as cause or effect; often the negligence entails nothing more than defendant providing erroneous information to the plaintiff.

The inquiry should begin with an examination of the nature of the emotional harm suffered, to bar recovery for grief and other non-compensable harms. The injury must be more than trivial and temporary. Plaintiff must prove these points.

This article focuses on what must be proved at trial, and how. Plaintiff must show by clear and convincing evidence that defendant's act(s) are the actual proximate cause(s) of the harm suffered, and are more than mere triggers, more than contributing factors, and more than the first event in an attenuated chain.

Defendant may present affirmative defenses, which, if proved by a preponderance of evidence, will relieve him of liability: 1) The injury would not have occurred to a normal plaintiff of average temperament (not one with "thin skin"); 2) plaintiff's conduct was an important contributing factor, either because he voluntarily placed himself in a precarious position, failed to mitigate the harm, or took some other inappropriate action.

BACKGROUND

This note discusses the development of negligently inflicted emotional harm as an independent tort, a relatively new cause of action. Traditionally, the common law was reluctant to recognize emotional harm and allowed recovery for it in two broad
categories: where the harm was intentionally, rather than negligently, inflicted, and where there was some independent cause of action to which the emotional harm claim was appended and thus said to be "parasitic." Examples of the "host" or primary causes of action are breach of contract and personal (physical) injury caused by defendant's negligent conduct.

Since about 1970, courts in a small number of states have decided that negligently inflicted emotional harm is an independent cause of action. The pronouncements usually take the form of a decision overturning the precedent which allowed the claim only when it was parasitic. The majority of states today require physical injury to the plaintiff or to a victim whose injury is witnessed by a plaintiff-bystander.

Many commentators have applauded and encouraged this movement, seeing it as long overdue. To a large extent, the movement may be seen as recognizing that the reasons which courts gave in the past for barring recovery in non-parasitic cases are either invalid or at least less compelling today. To begin our inquiry, we must review these reasons.

Probably the single greatest concern was that a plaintiff would get undeserved compensation through fraud. This could be the result of the difficulty in objectively confirming a plaintiff's allegations that he had indeed suffered an emotional injury. Even where the fact of the injury was not questioned, its magnitude was still uncertain. Perhaps the plaintiff was exaggerating, and it was impossible to prove otherwise. In the final analysis, the court or jury would have to rely almost exclusively on the plaintiff's testimony.

If the injury were real, the next difficulty was proving the causal link between defendant's conduct and plaintiff's injury. There might be a time lag between

---

1 Even at the turn of the twentieth century, two kinds of cases were exceptions to the general rule of disallowing recovery for emotional harm: cases involving the negligent mishandling of corpses, and cases involving erroneous notification by telegram of an event such as the death of a close relative.

Prosser and Keeton on Torts, 5th ed., at 362, point out the rationale for the exception: "What all of these cases appear to have in common is an especial likelihood of genuine and serious mental distress, arising from the special circumstances, which serves as a guarantee that the claim is not spurious."

Modern counterparts to these exceptions are discussed briefly infra.

2 It is useful to distinguish the various kinds of physical injury which pertain to emotional harm cases. The first kind is physical injury to a plaintiff directly caused by a defendant's negligence; this physical injury may be the cause of subsequent emotional harm, or there may be another cause. In either event, the resultant emotional harm is considered parasitic, and is addressed herein only in a cursory fashion. We are primarily interested in non-parasitic, "independent" causes of action. The second kind of physical injury is the result of the emotional harm. As we shall see, many courts require a plaintiff to prove physical manifestation of emotional harm.

Yet another kind of physical injury is that suffered by a third party victim, the injury witnessed by the bystander-plaintiff. In most cases, injury must be actual and serious for recovery to be considered.

or alternative explanations. An especially perplexing question, not resolved today by any means, was distinguishing between an injury proximately caused by a defendant's conduct--one where the plaintiff deserves recovery--and an injury in fact caused or explained by an unusual or abnormal or undue susceptibility to harm on the part of the plaintiff; this is not the kind in which recovery should be allowed.\footnote{This question is addressed infra, nn. 46-61 and accompanying text.}

Although hard to resolve in a single case, these concerns were not rejected solely because of their difficulty. Courts feared that once the rules were relaxed there would be a flood of litigation, and even trivial claims would have to be heard. Among other concerns was the belief that allowing recovery would place undue responsibility on defendants. Liability could be extended beyond the defendant's culpability in at least two ways.

First, if the barriers were removed, a defendant might become liable for significant injury where very clearly he was unlucky enough to have drawn a plaintiff whose pre-existing susceptibility to harm was, if not the very cause of the injury itself, the cause of the severity of the injury. The "thin-skull" plaintiff rule in physical injury cases held that "you take your plaintiff as you find him." That is to say, once defendant is negligent, he is liable for the full extent of damages, even if no reasonable defendant could have foreseen the magnitude of the injury. Should this concept be extended to the arena of emotional harm, where there are already difficulties enough?

The second was the possibility that once defendants were found liable for a new cause of action, plaintiffs would multiply not only in general, but for the incident at hand. For example, if the mother of a child killed in an auto accident could recover for emotional harm, what about father and grandmother and aunt and second cousin? Where would lines be drawn?

These concerns were the product of their time, reflecting an era in which the study of mental health, psychology and psychiatry were in their infancies, and the older fields of medical science were very much limited in diagnostic capability compared with today. Also, personal liability insurance was not as common or prominent as today.

It is perhaps an oversimplification to say that the judicial system simply did not have the means, fortitude or experience to solve the problem; by "the problem," I mean the difficulty in establishing principled methods to distinguish deserving from undeserving plaintiffs and claims. Two questions arise from all of this: 1) Given the concerns cited, how did the courts respond? 2) Are the concerns expressed above still valid?

ALLOWING RECOVERY UNDER ARBITRARY RULES

The answer to the courts' reaction is that they grudgingly gave ground when faced with a particularly compelling set of circumstances, and granted plaintiff at...
least a chance to present his case. From today's perspective, it appears that the fear of allowing recovery to an undeserving plaintiff was stronger than the fear of an unjust denial to a deserving one. A paramount concern remained a guarantee that there truly was an injury. Thus, where facts were similar but perhaps not quite as compelling, as cases were distinguished, a rule evolved which only later was seen to be unfortunately rather arbitrary.

As we examine this, our purpose is not to trace the history of arbitrary rules, which has been done elsewhere, but rather to demonstrate in what way they are arbitrary. At some point, the circumstances of a case would drive a court to allow recovery for serious fright suffered by a plaintiff through defendant's negligence. The fear might be of impending physical harm. One way to increase the certainty of genuineness in such a situation is to require that there be physical impact between the object feared and the plaintiff. Very probably, the judge who first thought of this impact rule envisioned a case (or probably it was thrust before him) where the defendant's negligence created some physical movement, perhaps at high speed, threatening collision with plaintiff. The impact requirement does serve as a very crude filter; it can separate a valid claim from one brought by another plaintiff in the same situation, but who was at a great enough distance from the threatened collision - let us say several hundred feet - so as to make groundless his fear and attendant emotional injury.

But the problem is that the requirement of physical contact became an all-or-nothing rule. Any contact would do, no matter how slight. (It is a little difficult to see what else could be done, once the rule was established.) However, the problem is that there is little difference between Plaintiff A, who fears an impending high-speed collision and whose arm is lightly brushed by the passing vehicle, and Plaintiff B, who fears the same collision but whom the vehicle misses by a hair's breadth. More precisely, the legal distinction between the plaintiffs is out of proportion to the virtually insignificant, and certainly meaningless, distinction between what happened to them physically. Such is the nature of the arbitrary rule.

Where did all of this lead? Until about 1970, courts did not recognize emotional harm as an independent tort, and required in most cases a showing of physical harm - that is, an objectively demonstrable ailment directly caused by emotional distress. Starting at this time, a few states did extend liability to independent emotional harm.
claims but even today, the list is short, and some partial retractions have occurred in forefront states.

A MODERN EVALUATION

Undoubtedly, the new trend was in large part based on the perception of several courts that the old fears were no longer valid, had dramatically reduced importance, or were outweighed by considerations of justice. Let us see how perceptions shifted, and then evaluate whether any of what I have referred to as the old concerns still merit our consideration.

Probably the single greatest difference in perception involved a steady growth in the courts' confidence that medical science could (1) establish that emotional harm had occurred, and (2) identify the causative factor(s). The development of that part of medical science dealing with the psyche had lagged behind the "hard" sciences such as biochemistry for a long time, but eventually flowered and was accepted. It seemed that tort law had not changed nearly as fast as, to pick one discipline as an example, psychiatry; it was argued that it was time for the law to catch up with modern capabilities.


9 This is extensively discussed in Miller, supra note 6.

10 A fairly comprehensive analysis of the old concerns, applied to a bystander case, in which recovery was allowed, appears in Sinn v. Burd, 486 Pa. 146, 158-64, 404 A.2d 672, 678-685 (1979). Some of the six notes infra appear as notes in that case.

11 Id. at 158-60, 404 A.2d at 678. "A survey of cases from other jurisdictions does not show the development of a logical and consistent rule, but reveals that the trend is a hesitant abandonment of...artificial restrictions and barriers to recovery in favor of a greater reliance on general tort law principles and the contemporary sophistication of the medical profession to test the veracity of claims for relief." Leong v. Takasaki, 55 Haw. 398, __ 520 P.2d 758, 762 (1974).

12 Courts "have recognized the maturation of medicine" in cases where emotional distress was intentionally inflicted, or where claim was parasitic, but "where the defendant is less culpable...courts have been reluctant to recognize the enhanced ability of science to measure mental distress." Comment, Recent Developments, Torts, 63 GEO. L. J. 1179, 1184-5 n.30 (1975).
The net result of heightened confidence in medical science was this: If it could be confirmed that there was harm - in the absence of physical injury - it was no longer necessary for the emotional harm to be parasitic to a separate cause of action to ensure genuineness.\(^{13}\)

What about the other "policy" reasons which had prevailed in the past? While advances in medical science were seen as deterrents to fraud, it was also argued now that the mere possibility of fraud was hardly a reason to deny plaintiffs their day in court.\(^{14}\) The concern over a flood of litigation was largely abandoned, even at times by courts holding the old line.\(^{15}\)

Finally, the concern of "unduly" extending liability without limits was seen as a problem which could be controlled. And, as one commentator pointed out, it seemed more sensible to draw a line other than at the point of an absolute bar to recovery.\(^{16}\)

My own perspective of this situation is that much of the previous concern is gone, but what remains is still important. Less danger exists today that a completely fraudulent claim will be brought, and the litigation flood is largely imaginary; further, trivial claims are *de facto* weeded out by the existing legal structure and process. There still remain serious issues regarding an undue extension of liability in cases where defendant is culpable of something, and juries tend to sympathize with an injured plaintiff, and "tack on" an emotional harm award, if they get the chance.

The acceptance of a completely independent status for the tort of emotional harm is well established in only a minority of jurisdictions. Where the claim is not

\(^{13}\) "Because other standards exist to test the authenticity of plaintiff's claim for relief, the requirement of resulting physical injury, like the requirement of physical impact, should not stand as another artificial bar to recovery, but merely be admissible as evidence of the degree of mental or emotional distress suffered." *Leong*, 55 Haw. at 520 P.2d at 762.

\(^{14}\) "Any rule which seeks to bar fraud...by withholding legal protection from all claims, just and unjust, employs a medieval technique which...is scarcely in keeping with the acknowledged function of a modern legal system." *Bystander's Recovery for Negligently Inflicted Mental Distress*, 29 ARK. L. REV. 562, 564-65 (1976) (quoting *Leflar & Sanders, Mental Suffering and Its Consequences - Arkansas Law*, 7 U. ARK. L. SCH. BUL. 43, 60 (1939)).

\(^{15}\) "[W]e point out that courts are responsible for dealing with cases on their merits, whether there be few suits or many; the existence of a multitude of claims merely shows society's pressing need for legal redress." *Dillon v. Legg*, 68 Cal.2d 728, 733, 441 P.2d 912, 917 n.3, 69 Cal. Rptr. 72, 77 (1968) (allowing recovery to a bystander).

The contemporaneous and opposed case is *Tobin v. Grossman*, 24 N.Y.2d 609, 249 N.E.2d 419, 301 N.Y.S.2d 554 (1969). "This court has rejected as a ground for denying a cause of action that there will be a proliferation of claims. It suffices that if a cognizable wrong has been committed that there must be a remedy, whatever the burden of the courts." *Id.* at 615, 249 N.E.2d at 422, 69 Cal. Rptr. at 77.

\(^{16}\) "If a line of circumscription is to be drawn for the sake of public policy, or even in the application of traditional tort principles, is it not more reasonable and humane to draw it somewhere other than at the point where no recovery is allowed simply because drawing the line elsewhere is difficult? Is not the line drawn in...*Dillon* more reasonably arbitrary than that drawn in *Tobin*?" *Simons, Psychic Injury and the Bystander: The Transcontinental Dispute Between California and New York*, 51 ST. JOHN'S L. REV. 1, 21 (1976).
appended to a "host claim", the necessity to show physical injury in order to enter the courthouse is logically the remainder of an era where other arbitrary rules were used to separate plaintiffs. It is much more principled to dispense with these rules altogether and to examine the problem from a different viewpoint.

PROPOSED APPROACH TO AN EMOTIONAL HARM CASE

The viewpoint to consider is the course of a trial which may be likened to an obstacle course, with recovery of damages as the prize, and with several discrete hurdles to be overcome by the plaintiff. Then, before the finish line, the defendant will have an opportunity to disqualify the plaintiff because of the latter's status or conduct.

More specifically, the plaintiff will be required to do the following: 1) Prove by a preponderance of the evidence that the defendant in fact committed the acts of negligence which the plaintiff alleges are the cause of his injury. Obviously, this imposes no new burden on the plaintiff; 2) demonstrate that the injury is genuine, serious, and clearly not primarily caused by any incidental circumstance which the defendant could either control or foresee. Further, the injury must have been caused by something other than a stimulus commonly encountered in human experience; 3) prove by a heightened standard - clear and convincing evidence - that defendant's acts caused the harm. The actions must be more than a contributing factor, more than the first event in an attenuated chain of causation, and more than a mere triggering event. The standard of clear and convincing proposed here is not novel, but is not what is commonly required.

The defendant will have the opportunity to present affirmative defenses; if these are proved by the preponderance of the evidence, they will bar recovery by the plaintiff. Either of the following defenses would be fatal to the plaintiff's recovery: 1) A showing that the injury was caused by an unusual susceptibility of the plaintiff, or alternatively, that a normal plaintiff would have suffered so little injury as to render it non-compensable. 2) A showing that some aspect of the plaintiff's conduct was inappropriate - he either voluntarily put himself in an exposed position, he failed to do anything to mitigate or avoid the injury, or there is some act or response by the plaintiff which is inconsistent with the later claim of injury.

SOME GENERAL CONSIDERATIONS ABOUT HARM

The first requirement listed as an element of the plaintiff's case is proof that the acts alleged to have caused the injury were actually committed in fact by the defendant. This is not controversial and is of course an element of every tort claim brought by a plaintiff.17

17 Plaintiff may, in the alternative impute responsibility to the defendant via vicarious liability. That is, under limited circumstances, a plaintiff injured by a child may impute liability to the parent as the defendant. Thus, the defendant may be liable because of the act of another.
However, once we consider the nature of the harm suffered, we begin to encounter problems. The plaintiff should have the burden to prove that the harm is genuine, serious, and clearly not primarily caused by any incidental circumstance which the defendant could either control or foresee. The requirements of "genuineness" and seriousness may or may not coincide. And unavoidably intermixed is another concept which is closely related: the harm must be of a kind that we choose to compensate; it must be the "right kind" of harm.

DISTINGUISHING HARM FROM INJURY

No small part of a generalized and historic opposition to allowing recovery for emotional harm has been the difficulty of identifying exactly what injury is involved. One does not identify the harm by listing its symptoms; saying that the harm is sleeplessness, or headaches, or uncontrolled crying does little more than confirm that there is a manifestation of injury. At the risk of oversimplification, the answer may be labeled "primary effects" (grief, shock, fear or anger). These primary effects then may or may not give rise to some manifestation of physical or psychic harm which can be confirmed objectively.

This concept may be understood by consideration of a typical "bystander case," in which we see a sequential chain of causation as follows:

Negligent conduct
causes
Physical injury to victim
causes
Primary effect (e.g., shock) on plaintiff
causes
Manifestation of injury in plaintiff

One should distinguish between the manifested injury and its cause - in the example above, shock - because in many cases, only some of the plaintiff's injury is caused by the defendant's conduct. Real life is full of situations involving a complex set of factors, and it is not always possible to identify a single cause.

When we ask what the injury or harm is that is being compensated, only the generalized type of distress such as shock or fear should qualify. The question is: did defendant's action cause shock, rather than: did defendant's action cause sleeplessness. If the act caused the primary distress, that is enough. More will be said on the second link -- whether the shock caused the sleeplessness - in connection with plaintiff susceptibility.

AN EXAMPLE OF AN UNCOMPENSATED HARM: GRIEF

A benefit of this analysis is that it allows us at the outset to differentiate types of harm for which there is generally no relief. The most important of these is grief, by which I mean that profound sorrow occasioned by one's awareness of a victim's loss of life, or perhaps a devastating physical injury. Beyond grief, it becomes more difficult
to avoid value judgments, but one commentator has listed for exclusion from recovery "mere' upset, dismay, humiliation...and anger".\footnote{Comment, \textit{Negligence and the Infliction of Emotional Harm: A Reappraisal of the Nervous Shock Cases}, 35 U. CHI. L. REV. 512, 517 (1968).}

Fright seems to be somewhere in between, perhaps because it may be profound and justified, or in contrast, trivial and short-lived.

Probably no modern case in the field of negligent infliction of emotional harm has received more attention than \textit{Dillon v. Legg}.\footnote{68 Cal. 2d 728, 441 P.2d 912, 69 Cal. Rptr. 72 (1968).} A consideration of \textit{Dillon} reveals important aspects of grief as it typifies the kind of harm where no recovery is provided. This 1968 California decision set forth three guidelines which it stressed were to be used as tests of foreseeability by courts in their deliberations on bystander cases. \textit{Dillon} proposed that the bystander be allowed to recover if three conditions were met:\footnote{Id. at 36, 441 P.2d at 920, 69 Cal. Rptr. at 80.} 1) The victim and the bystander plaintiff must be closely related; 2) the bystander was nearby; and 3) the bystander's injury was a direct result of a sensory and contemporaneous observance of the accident.

Note carefully the last factor, which has the function of distinguishing the harm of shock from the harm of grief. Consider two accidents in which a child is killed. In the first, the parent is far removed and learns of the tragedy only later. In the second, the parent is a contemporaneous witness. The former is not entitled to damages for emotional harm while the latter, the bystander, recovers. Why? Because the compensation is not for the grief at the child's loss, but rather for the shock of witnessing the accident. When we say freedom from emotional injury is to be a protected right, we do not mean to compensate for grief. Any number of decisions following \textit{Dillon} may be cited on this point.\footnote{Cortez v. Macias, 110 Cal. App. 3d 640, 167 Cal. Rptr. 905 (1980) is one example. Here, a mother's claim for emotional distress following the death of her young child after alleged physician malpractice is not allowed as a matter of law at trial, and appellate court upholds. In this poignant case, a mother leaves febrile child in the emergency room to settle the bill and upon return is told of the child's death. All of the plaintiff's manifestations of emotional harm follow this notification, thus showing that there was no contemporaneous observance of a shock-inducing event. The harm is clearly grief at the death of the child rather than shock induced by witnessing any negligent act of the defendant.}

\section*{OTHER HARMs AND ARBITRARY RULEs}

Of course, other rules have in the past been used to differentiate types of harm. Certainly, the foundation of the "zone-of-danger rule" is a willingness to compensate when the harm entailed is fright, or fear for one's physical safety. The theory is that recovery is proper where the defendant's negligence exposes the plaintiff to serious bodily harm, and recovery should not depend on whether impact occurred. The difficulty is that the zone of danger is a physical concept and may not necessarily
correspond to the individual's perception of fear. That is, one person may perceive danger at a given distance, while another does not. The zone rule may perhaps be defended in that it works most of the time, but very likely, it works by denying some deserving plaintiffs at the expense of ensuring that undeserving ones are barred from recovery.

SERIOUSNESS OF THE HARM

The desire to limit recovery to cases where the injury is serious is historic.\(^22\) When we say that seriousness should be a requirement before an injury qualifies to be compensated, we may mean seriousness in one of several different aspects. Do we mean the harm must be of a serious nature? Do we mean that the intensity or magnitude of the harm must be great enough? Does the duration of the harm matter?\(^23\)

Taking these in turn, we have seen in the preceding section that the nature of the harm matters, and grief, despite its seriousness, is not compensated. While it is difficult to generalize, there should not be compensation for feelings of embarrassment or the fear of ridicule even where these may be traced to a negligent act of defendant.

In *Woodell v. Pinehurst Surgical Clinic*,\(^24\) plaintiff was erroneously and negligently told by defendant clinic that she would deliver twins. After a single child was delivered, the mother brought a claim for several different injuries, only the emotional harm claim concerning us here. The basis of the claim was humiliation and embarrassment, and the trial court's dismissal was upheld because the plaintiff suffered no physical injury.

This is really not a satisfactory way to settle this case. As we have seen, at least some jurisdictions no longer require physical injury, and certainly, there may well be serious harm in its absence. The case ought to be dismissed, but on the grounds that the injury is not actionable because it is trivial, its nature is common, and it is at least partly either imagined by the plaintiff, or contributed to in a large way by her. Here the plaintiff's real injury is disappointment; the plaintiff looked forward to having twins and likely proclaimed her expectation to the world at large.

\(^{22}\) Among the cases stipulating seriousness as a necessity is *Paugh v. Hanks*, 6 Ohio St. 3d 72, 451 N.E.2d 759 (1983), but it includes the comment, "Finally, we note that determination of both seriousness and reasonable foreseeability must be accomplished on a case-by-case basis." *Id.* at _, 451 N.E.2d at 767. *Paugh* is an interesting bystander-like case and the analysis follows *Dillon* to a large extent despite the fact that there is no identifiable victim! The nearest substitute to a victim is plaintiff's house or fence.


\(^{23}\) See text, infra.

DEGREE OF SEVERITY

Even if the harm is one we want to compensate because it is serious, what if the extent of the injury is minor? This is not a very fruitful area for inquiry, because questions of degree are ultimately questions of fact, and are for juries to determine. But let us at least define the problem. Consider a fact pattern for a bystander injury case, including the death of the victim, and for the sake of argument, a bystander meeting the Dillon guidelines. Thus, we are ready to allow the case to proceed, primarily for the purpose of verifying the factual allegations of the plaintiff. Now, let us make this variation - the injury is minor, and the bystander correctly perceives this. 25 Perhaps the victim is grazed, and requires two stitches at a doctor's office. If this is the case, we would not expect a normal bystander to become an emotional harm plaintiff - she is very unlikely to succeed, the reward would certainly be small, and she may have a lot of difficulty in getting a lawyer to take her case.

DURATION

If the harm is short-lived, it may not be serious. Even if we can get universal agreement that this is a valid principle, we are then faced with deciding what length of time is "enough." Usually, this needs to be determined in the framework of the circumstances of a case. The problem does arise in what I call a pure misinformation case, which has these characteristics: 1) There is no harm other than the victim having an erroneous perception of reality, said perception resulting directly from a communication from the defendant. As an example, a medical misdiagnosis could be merely misinformation if the course of treatment was not affected by the misdiagnosis.

25 But what happens where the bystander's perception is erroneous? That is, suppose the physical injury to victim is not serious, but bystander thinks it is? We can examine two bystander cases which address this issue.

In Barnhill v. Davis, 300 N.W.2d 104 (Iowa 1981), the plaintiff drives through an intersection and waits for his mother, who is in a following car. Defendant's car negligently strikes the mother's car, and she receives very minor injuries. Nonetheless, the plaintiff later sues the defendant for negligently inflicted emotional harm, alleging sleeplessness and pain in his back and legs; a doctor testifies that these symptoms were caused by plaintiff witnessing the accident and fearing for his mother's safety. On appeal, the court reverses the trial court and allows the plaintiff to go to trial for factual determination, holding that it is enough if a reasonable man would believe, and the plaintiff at bar did believe that there was a strong possibility of serious injury to the victim.

In contrast, Ramirez v. Armstrong, 100 N.M 538, 673 P.2d 822 (1983), holds that the accident being witnessed must result in serious physical injury or death to the victim in order for the bystander to recover. The difference in the holdings is interesting because it points out the deficiencies in the Barnhill holding. Implicit in Barnhill's holding is the idea that recovery may be had for emotional distress independent of whether it is "justified." A further "explanation" of the holding is that the plaintiff is allowed to recover for the time in which there was uncertainty as to the victim's fate. Curiously, if that is the reason, the plaintiff would recover a greater amount by never inquiring into the actual physical condition of the victim, certainly an undesirable and inconsistent outcome.
(this happens when the misdiagnosis is corrected promptly); 2) any actions taken by
the victim based on the misinformation are not in themselves harmful.

We must observe that a prerequisite of a claim based on misinformation is the
correction of the plaintiff's perception. Usually, the defendant checks and advises the
plaintiff of the true situation. However, often plaintiff is actively engaged in the
process of error correction. In either case, plaintiff's claim is logically limited to harm
suffered during that period of time when he was under the erroneous perception.

Whether some threshold limit has been passed could be a jury question, and this
presumes that the concept of the threshold is part of the court's jury instruction.

---

26 It has been suggested that plaintiff in a libel case be barred from recovery where
defendant retracts in a reasonable time. A retraction published as broadly as the libel
repairs much of the harm. See LeBel, Reforming the Tort of Defamation: An
Accommodation of the Competing Interests Within the Current Constitutional Framework, 66
NEBR. L. REV. 249, 304-315 (1987). A misinformation case is similar in that a
prompt correction reduces harm; it seems evident that the reduction in harm is directly
proportional to the speed of the correction. If we examine the extreme case of a
virtually simultaneous, yet sequential misinformation and correction, there is no harm.
Imagine Kaiser Foundation Hospital telling the Moliens (see text, infra) that Mrs. Molien
had syphilis, and five seconds later retracting in an authoritative manner. No jury
could reasonably find that a subsequent divorce had as a proximate cause the
misdiagnosis which existed as a fact in the minds of the couple for a period of five
seconds. But, of course, the 5-second example is unrealistic; the difficulty is that a
"real-world" problem is going to be one of degree.

In one respect, there is a stronger argument for barring recovery in the case of
rapid correction of misinformation than for barring libel recovery after a retraction.
In correcting erroneous non-libelous misinformation provided to one person, it is certain
that the correction gets to the victim; in libel retraction, where the harm is caused by
misinformation remaining in the minds of perhaps millions of readers, there is no
assured way to reach every one.

Of course, libel and misinformation (as the latter term is used herein) are far
from identical; plaintiff/victim in what we have been calling misinformation cases does
not need a retraction to be published - his injury is not dependent on the belief of third
parties.

27 Martell v. St. Charles Hospital, 137 Misc.2d 980, 523 N.Y.S.2d 342 (1987) is decided
contra my position. Defendant's error is corrected within two days, but the court holds,
"...the recovery...should not be artificially limited simply because in the opinion of the
defendant...the plaintiff's suffering should have ended at a particular moment or upon
the happening of a particular event." Id. at 990, 523 N.Y.S.2d at 352.

At a further point in the same opinion, "If the defendants...are responsible for
the injury the damages are not to be limited merely because this particular plaintiff was
susceptible to greater damage than might be another individual." Id. Because of the
phraseology, there is room for interpretation, but this last quotation could be read as
extending liability in cases where aberrant susceptibility of the plaintiff is the
dominant cause of the injury; this is very much a minority position.

28 As an alternative, it is not beyond the pale of reason that an arbitrary timeframe
or set of them be used. How might this work? Let us confine our inquiry to three
categories of misinformation, and suggest some limits for consideration. In each case,
there would be no recovery if the error is corrected by defendant within the limit
stipulated, so long as defendant's culpability is simple negligence, as opposed to
recklessness: Information to victim that he has a life-threatening disease. (The
contrary, erroneous information that victim is free from a life-threatening disease, can
hardly be a "pure" misinformation case, because it is almost sure to entail omitted
treatment.) 48 hours. Information to the victim that he has a medical condition less
serious than life-threatening. One week. Information that a relative has died. 8 hours.

Difficulties can be seen immediately; it is hardly likely that a comprehensive
listing could be developed, even if a legislature or court were so inclined.

86
OTHER NON-ACTIONABLE HARMs

There are probably some special cases where the harm has some peculiar characteristic that renders it unappealing from the viewpoint of recovery. And, there is a peculiar kind of case in which we readily sympathize with the victim, but reason precludes recovery because it is difficult to find a true harm. These are what I call the "false good news" cases.

For example, suppose a person with a life-threatening malady who surprisingly and erroneously is told by a doctor or hospital that there has been some error, and the person need not fear that malady. If by the nature of the case, there is no deterioration of the victim's condition proximately caused by any act or omission stemming from the misinformation, it is hard to see why there should be recovery under an emotional harm claim, when the true state of affairs is re-established. Physically, the victim is back where he started, and certainly, he cannot claim harm for the intervening period when he thought he was healthy; during that period, he experienced the opposite of harm, albeit under false pretenses. The claim would seem to be based on an aggravation of an already distressed person facing death, where until the sequence of false hope/"correction," there was no one culpable and hence that person was not a victim in the sense of incurring an injury caused by another.

Kossel v. Superior Court comes close to the fact pattern and result suggested above, but the case is settled on other grounds. In Kossel, a cancer patient was told that there had been a mistake and that he did not have cancer. Soon thereafter, however, the "mistake" was cancelled and the patient was back where he started. It is unclear from the opinion whether the patient had any claim that the faulty diagnosis affected his medical treatment. If it did, that is malpractice, in which the harm is physical, and the injury can be objectively demonstrated. The claim in Kossel, however, was brought by the spouse rather than the patient. The relationship, or lack thereof, between the plaintiff-spouse, and the defendant hospital, is the ground for the dismissal at appeal; she could not recover because she was not the direct victim of the misinformation.

This seems to be the right result, but not for the right reason. The court's decision leaves room for a recovery by the "right kind" of plaintiff where there is no injury. The starting point ought to be whether there was an injury crossing some threshold of seriousness; this inquiry should precede any relational questions.

THE BURDEN OF PROOF ON CAUSATION

In analyzing emotional harm claims, we can consider the elements to be demonstrated by a plaintiff to overcome his case, and then go on to look at certain affirmative defenses offered, as a series of hurdles or barricades for the plaintiff to pass in order to prevail. Some of the barricades should be higher than others in order to meet the concerns of those opposed to recovery for emotional harm that still retain

---

some validity. One way to elevate a barrier is to require heavier burdens of proof for some elements of the plaintiff's case.

Every claim brought against a defendant for emotional harm rests on a discrete act or omission of the defendant. (For the sake of simplicity, I shall use the term "act" or "action" to include "omission" in the remainder of this paper.) Proving that this action did in fact occur is the first element in the case; no special burden seems appropriate here. If contested, the plaintiff has this burden, but the ordinary standard of proof - preponderance of the evidence - will suffice.

At the next step of the analysis, it is proposed that causation - that is, the link between the act and the harm - be shown by a clear and convincing standard. This proposal is not original; clear and convincing is the standard for recovery for negligently inflicted emotional harm in Virginia,30 for example. Imposing this standard has several important consequences.

First, it acts as a deterrent to some plaintiffs. They or their prospective attorneys may be less willing to enter the race where the burden to be met is higher. Secondly, for those who do proceed to trial, clearly a properly instructed jury will find it more difficult to find for the plaintiff. Thirdly, there will be an increase in the number of cases where judges will decide using their own discretion that a clear and convincing standard has not been met, and find that the plaintiff fails.31

With this preface, let us examine the key causation questions that must be satisfied by clear and convincing evidence: Is it clear that the defendant's act is significantly more than a contributing factor; if several factors can be identified, is the act in question the dominant causative event?

ATTENUATION

As an example of how the proposed causation standard would resolve a troublesome case better than the actual decision, consider Kelly v. Kokua Sales Supply, Ltd.32 An unfortunate accident occurs in Hawaii whereby defendants, through their negligence, are responsible for the deaths of a child and her mother. The tragedy is compounded when a short time later, the grandfather of the child victim (he is also the father of the other victim, but let us refer to him as grandfather here) is notified by telephone at his home in California and suffers a fatal heart attack. The decedent's estate brings a claim of emotional distress against those held negligent in the accident.


31 Anderson v. Liberty Lobby, 477 U.S. 242 (1986), held that where a clear and convincing standard of proof is required of plaintiff on a given element of his case, as a matter of law, the trial court must use this standard in response to a motion for summary dismissal. It is presumed here that the Liberty Lobby decision applies beyond defamation to a tort such as negligent infliction of emotional distress.

Let us set aside the question of whether the grandfather had a pre-existing condition, and the possibility that grief over the loss of the grandchildren rather than shock is the cause of the injury (it seems very difficult to decide either way since the decedent is not available for examination). It is more useful to examine the way in which the case is disposed of by the Hawaiian court.

The court does not consider this a bystander case. If it did and applied the Dillon factors: (1) close family relationship, (2) contemporaneous observance of the accident, (3) physical proximity, it is clear that only the first is met. The case arises, however, following the decision in Hawaii in Rodrigues v. State\textsuperscript{33} to recognize emotional harm as an independent tort and to decide cases based on general principles of tort law. Thus, nothing like the Dillon guidelines apply here. However, in what must be seen as a major retreat from the earlier Hawaiian decisions in Rodrigues and Leong v. Takasaki\textsuperscript{34} (which is a bystander case),\textsuperscript{36} the court in effect applies a Dillon guideline of physical proximity, limiting recovery in such cases to plaintiffs "located within a reasonable distance from the scene of the accident."\textsuperscript{35} Much emphasis is placed on the great geographical distance between California and Hawaii, and plaintiff loses.\textsuperscript{36}

The decision is confusing and inconsistent. Part of the confusion may be attributed to the backtracking phenomenon. How could the case have been handled without falling back to a Dillon-like guideline? It seems clear that the primary motivation of the court is the concern with extending the defendant's liability too far. But if a limit is desired, geography is hardly a reliable index (suppose the grandfather lived a block away - would that have made a real difference?) Much more appropriate in this case would be denying recovery because of the attenuated nature of the claimed causation. Defendant is clearly responsible for the deaths of the direct victims of the accidents, but based on foreseeability, he can hardly be responsible for consequences to relatives of the victims without limit, even where it can be proved that shock rather than grief is the "explanation" of that injury. Here, there is an implication that the telephone call and perhaps its abrupt nature played a key role. But, of course, the defendant had no part in making this call.

Plaintiff's case should fail because the grandfather is not a direct victim of the defendants' negligence - he had no contact with the defendant - nor is there any existing relationship between them giving rise to a duty which was breached. Nor could the grandfather be construed as a bystander. Grandfather was an innocent victim, and the defendant was culpable of something, but that is not enough. All of this is another way of saying that causation can only be attributed to the defendant in this case by an attenuated chain. If this is the focus, the disposition is straightforward.

\textsuperscript{33} 52 Haw. 156, 472 P.2d 509 (1970).
\textsuperscript{34} 55 Haw. 398, 520 P.2d 758 (1974)
\textsuperscript{35} Kelly, 56 Haw. at _, 532 P.2d at 676.
\textsuperscript{36} Actually, the summary judgment for the defendants granted by trial court is affirmed.
TRIGGERS

In *Kelly*, attenuation lies in the intervening telephone call, and in the physical remoteness of the grandfather from the accident caused by the defendants' negligence. In what way other than attenuation may the plaintiff fail to show causation? If the negligence of the defendant serves merely as a trigger, no proximate causation occurs.

Consider *Molien v. Kaiser Foundation Hospitals.* This is another landmark decision in California, rivaling *Dillon* in the attention it received both in the literature and from courts throughout the country. Valerie Molien, a hospital patient, was negligently misdiagnosed as having syphilis while undergoing some routine testing. Immediately thereafter, her husband Steven underwent a similar test and was found to be free of syphilis (presumably at this point the test on Valerie was repeated and the earlier test seen to be in error). Despite Steven's negative test result, Valerie accused Steven of infidelity and eventually they were divorced.

When Steven brought an emotional harm action against the defendant hospital, the trial court sustained defendant hospital's demurrer on the ground that the plaintiff could not state a cause of action, not having claimed any physical injury from the emotional harm. But on appeal, and this is the significance of *Molien*, the California Supreme Court reversed, thereby establishing emotional harm as an independent tort for the first time.

It does seem a very strange case for this to have happened, because the facts are so far from compelling. It must first be observed that the misdiagnosis does not cause any substantial direct harm from a medical viewpoint; any culpability of the hospital lies in incorrect information rather than incorrect treatment. Secondly, whereas Valerie Molien may have had good reason for suspecting her husband of infidelity during that short time before her husband tested negative and her own diagnosis was corrected, that particular reason for suspicion expired with the correction of the error. She may have had other good reasons for her suspicions of infidelity, but the point is

---


Numerous courts have cited *Molien*. Shepardizing in September, 1988 showed that 132 state or federal cases mention *Molien*. Of these, seventeen cases are cited to show that *Molien* was followed. Some of these extend the principles of *Molien* to cases other than negligent infliction of emotional harm; a particularly fascinating case is *Kately v. Wilkinson*, 148 Cal. App. 3d 576, 195 Cal. Rptr. 902 (1983), wherein the plaintiff-bystander is the owner of a boat and the victim is killed in a water-skiing accident. Defendant's liability is based not on negligence, but on products liability.

39 Valerie Molien was treated with massive doses of penicillin, which the court characterized as "unnecessary and painful". While no timeframe is mentioned in the opinion, it seems likely that this did not extend beyond a few days. In any event, what she suffered was physical pain, no emotional, and that injury was of course suffered by her rather than Steven, who is the plaintiff in the case we are discussing.
that the other reasons are not linked to the defendant. Thus, all the defendant did in this case was trigger a set of suspicions, not cause them.\textsuperscript{40}

There are really two different kinds of injury in \textit{Molien}. The first of these is the emotional harm caused by the misdiagnosis; a section of the opinion where the court is imputing defendant's responsibility to the plaintiff, who is Stephen, not Valerie, says: "It is easily predictable that an erroneous diagnosis of syphilis and its probable source would produce marital discord and resultant emotional distress to a married patient's spouse."\textsuperscript{41} But this injury was, or at least should have been, transient, because the misdiagnosis was corrected. The greater injury, and the one for which the plaintiff seeks compensation is the emotional distress as the marriage disintegrates. Part of the claim is for medical costs for counseling in an effort to save the marriage. While exact chronology is not provided in the appellate court opinion, it seems reasonable to presume that the counseling was at a point later than the time when the true diagnosis was made known.

Certainly the facts here raise suspicions of an unusual plaintiff susceptibility or pre-existing condition. Usually, as discussed elsewhere in this note, it is the susceptibility to harm, meaning that there is too little resistance in the plaintiff's makeup to be able to cope with some stressful event. But here, the real susceptibility almost surely resides in the weak marriage of the Moliens, which was susceptible to divorce. The divorce of such a weak marriage was the cause of Steven's emotional distress, and not the harm caused by the misinformation. Defendant's negligence is at best a contributing factor, not even necessarily a "but for" factor. Can the Moliens show to a clear and convincing standard that but for the negligence, they would still be married?

It would be straining to call any susceptibility to divorce unusual given contemporary divorce rates. But if the argument is made that the Moliens' susceptibility to divorce is average, does this not come close to saying they had roughly a one-in-two chance of staying married for even 10 years? If this is so, how can a negligent event outside the marriage be the legal cause of harm under a heightened standard? Further, allowing recovery for emotional harm caused by divorce has enormous potential for a flood of litigation.

In summary, the case should be thrown out as a matter of law because the defendant's negligence did nothing more than trigger a chain of events. The plaintiff

\textsuperscript{40} Plaintiff conceivably could argue that defendant's actions are the cause, because whereas Valerie's suspicions are irrational in the face of the correction, those are after all still her suspicions, rational or irrational. The response to this is that defendant could hardly foresee that plaintiff would form opinions or suspicions or ideas based on a diagnosis made by defendant, and continue to hold them when the diagnosis was changed (and observe the yes/no nature of the test - you either have syphilis or you don't, with no middle ground).

\textsuperscript{41} \textit{Molien}, 96 Cal. App. 3d at 473, 616 P.2d 813 at 817, 167 Cal. Rptr. at 835.
cannot show even by a preponderance of evidence that the defendant's negligence caused the injury.42

ONLY A CONTRIBUTING FACTOR

Where it can readily be seen that the emotional harm is caused by several different factors, a case can be made for barring liability against the defendant unless by clear and convincing evidence it is shown that the defendant's actions are the dominant cause of the harm.

This could be applied to the case of Hoard v. Shawnee Mission Medical Center,43 which is also discussed below from the viewpoint of plaintiff susceptibility. In Hoard, there are a number of events: 1) A teenage girl is critically injured in an automobile accident; 2) parents go to defendant hospital and are told erroneously of death of girl, who in fact is at another hospital; 3) victim survives, but has sustained serious and permanent damage; 4) after a time lapse, parent plaintiffs develop manifestations of emotional harm.

Even a cursory review reveals the obvious fact that the plaintiffs have understandable reasons for emotional distress. However, the defendant hospital is tied only to one reason. Can the plaintiffs hold defendant liable for their harm? They should be required to prove by the heightened standard that the defendant's actions are the dominant causative factor.

AFFIRMATIVE DEFENSES

Defendant may present affirmative defenses which cancel liability if proved by a preponderance of evidence.

PLAINTIFF SUSCEPTIBILITY

If defendant can show by a preponderance of the evidence that the injury would not have occurred to a normal plaintiff, he should be relieved of liability. Alternatively, he may show that a normal plaintiff would have been injured only to a trivial degree. "Normal" means a person of average constitution, free of any aberrational emotional susceptibility, capable of coping with at least the stresses of daily life.44

42 The court instead concentrated on these questions: Did defendant owe a duty to this plaintiff? (Yes) May plaintiff recover even though he suffered no physical injury? (Yes)


44 See Simons, supra note 16, at n. 118, where it is contended that the "normal individual is impossible to define and consequently unavailable as a standard for recovery."
It is beyond dispute that reactions to stimuli and events vary with the individual perceiving them; this is true of physical reactions, and of psychic or emotional response. Our legal system has always been aware of this and it is an important factor in explaining why arbitrary rules were and are used to deny recovery. It is just too difficult to attribute or allocate causation to the defendant's actions and the plaintiff's heightened susceptibility to harm. The problem is by its nature intractable and resists attempts to resolve it by statistical data, for instance. One is never quite sure what the true cause is, or what proportion of causation lies with each factor.

An early attempt at analysis was made by Smith, who examined over 300 cases and who concluded that some 60% of them involved unusual plaintiff susceptibility. Little support is provided, and one gains the impression that Smith merely examined reports of cases, rather than made any independent examination. He also buttressed his argument with the observation that there is a 5/1 female/male ratio in the class of plaintiffs; this is corroborated if we accept the added inference that women are more prone to heightened emotional susceptibility than are men. Smith's work has been mentioned by other commentators and courts but it has not to my knowledge been confirmed or refuted by any other published work which includes a compilation of data or a statistical analysis.

Perhaps it is unnecessary to quantify this matter, because we can readily show that the plaintiff's unusual susceptibility is the dominant cause of the harm, at least in some cases. Let us examine some actual fact patterns. In *Williamson v. Bennett*, the plaintiff automobile driver was involved in a collision caused by the defendant's

---


47 See supra, note 4.

48 This work is plainly dated and based primarily on the author's interpretation of writings available to him on a large number of old cases. A perusal of his Appendix A reveals that many of the cases "investigated" are more than 30 years old. Thus, there is little direct research. Despite this weakness, the Smith study is cited by many modern commentators because, probably, there is no modern counterpart. Note, *Relation of Emotions to Injury and Disease: Legal Liability for Psychic Stimuli*, 30 VA. L. REV. 193, 306-7 (1943).

49 Id. at 280-1.


51 See, e.g., Sinn v. Bird, 486 Pa. 146, 404 A.2d 672 (1979); Amaya v. Home Ice, Fuel & Supply Co., 59 Cal.2d 295, 379 P.2d 513, 29 Cal. Rptr. 33 (1963) (the latter was a California bystander case, which set the standard until overruled by *Dillon*).
negligence. Damage to the vehicles was slight, and Mrs. Williamson was uninjured. After a few days, she developed serious manifestations of emotional distress, and attributed these to her vivid recollection of a recent accident involving a relative. Her relative was driving a car that was struck by a child on a bicycle, and the child was killed. Mrs. Williamson had not witnessed this accident, but said that at the time of her own collision (in which no bicycle was involved), she feared that it was an identical situation; she even went so far as to say that she was afraid to look at that portion of her car that was damaged for fear of seeing the (imagined) dead child.

The court on appeal found that Mrs. Williamson's susceptibility was abnormal, and further added that she would be barred from recovery based on fright, because her fear was for another, not herself. This case is one where we can readily see that a foreseeability analysis works. Defendant can foresee physical damage as a consequence of his negligence, and perhaps a possibility of some amorphous emotional upset, likely to be short-lived and not very serious. He could hardly expect the aberrational fears of this particular defendant, and the subsequent degree of injury. The case is chosen to show that the defendant here could easily and convincingly have shown that the bizarre response admitted by the plaintiff is the cause of her harm; such a showing by itself should rightfully serve to relieve the defendant of liability.

Hoard v. Shawnee Mission Medical Center is a misinformation case. The medical center negligently and erroneously told the Hoards that their 17-year-old daughter had died of injuries sustained in an auto crash shortly after being brought to the center. Actually, the girl had been taken to another hospital; the dead girl at the Shawnee institution was another victim of the same crash. Lisa Hoard, the accident victim, was still alive but had suffered extremely severe injuries and was comatose for six weeks immediately thereafter. Her recovery was limited, and she remained a considerable burden on her parents.

Many of the symptoms complained of by the parents in their action were not immediate, a factor in the court's decision to deny them relief. The court also obviously struggled with the difficulty in separating the causative factors, i.e., reaction to the misinformation and the accident itself.

---

52 251 N.C. 498, 112 S.E.2d 48 (1960).

53 This latter point is a mechanism the court uses to deny recovery on what I call a secondary ground. That is, the court says that recovery which might be allowed to someone in the zone of danger is based on the danger and fear to the plaintiff herself. In the Williamson case, plaintiff feared for some supposed party she merely imagined to be present. The primary point is not who she fears for, but the completely unexpected nature of her fear.


55 Id. at __, 662 P.2d at 1220-1.
to the physical injury and its resultant burden versus the aftermath of the shock induced by the negligent and erroneous notification of death.\textsuperscript{56}

But predisposition or heightened susceptibility was also a very important factor. In the case of Mr. Hoard, he had a prior history of cardiac problems, and some part of the post-accident harm seemed related to this condition, in the court's analysis. Mrs. Hoard had been taking medication for many years for nervousness; she had been earlier treated for depression, alcoholism and esophagitis; the latter was also an ailment she complained of after the accident.\textsuperscript{57} The court found as a matter of law that the causation alleged by the Hoards was too speculative to merit recovery from defendant.\textsuperscript{58}

**Plaintiff Conduct as a Bar to Recovery**

In addition to susceptibility, an affirmative showing that the plaintiff's conduct was an important factor should bar his recovery. What is intended here is an analog to contributory negligence in a manner which might be referred to as an "all-or-nothing" decision. Prior to more recent comparative negligence schemes characterized by reduction in the amount of the plaintiff's award, any finding of contributory negligence totally barred recovery.

This may seem draconian, but in practice I believe it would be applied only in clear-cut situations. A first example would apply to what we have previously defined as a pure misinformation case. Where defendant can prove that the plaintiff expressed serious disbelief upon the receipt of the information, and this was not merely an emotional reaction or a wish fulfillment, there can be no recovery, as a matter of law. That is, the jury never gets to say, "Plaintiff did not believe it, but we find for him anyway."

Such a rule would work best in the situation where the plaintiff's doubt is strong enough so that he seeks confirmation of the information from another source, and this search can be objectively demonstrated. In *Johnson v. State of New York*,\textsuperscript{59} the plaintiff is negligently and erroneously told by a hospital that her mother, a patient of the

\textsuperscript{56} Note that Kansas did not subscribe to the general exception remarked in the text, *supra* note 1, allowing recovery for emotional distress caused by erroneous death notification by telegraph. If they had, it might have made an important difference in this case. *Id.* at , 662 P.2d at 1220.

\textsuperscript{57} *Id.* at _, 662 P.2d at 1218.

\textsuperscript{58} "The alleged resulting damages are simply too conjectural, speculative, and remote in time from the incident to form a sound basis for measurement, and therefore no recovery can be allowed upon the basis of negligent infliction of emotional distress. The foregoing issue presented a question of law for the trial court and therefore a motion for summary judgment was properly sustained." *Id.* at _, 662 P.2d at 1222-3.

hospital, has died. 60 The burden of the case in the eyes of the court is to hold that there may be recovery without a showing of physical harm.

After the notification, the plaintiff and a relative express a definite skepticism about the information, and make inquiries which result in finding the error. 61 The opinion does not specify the time lapse between error and correction, but it could not have been very long, as the correction involved viewing the remains of the deceased at a mortuary. But that is only one point. What is of interest here is the profound inconsistency between (a) the alleged harm upon hearing the news and (b) the demonstrable disbelief of plaintiff. How can the news have been the cause of the harm if the plaintiff did not believe it? It seems that the real driving force for plaintiff's claim is not harm but indignation.

The rule proposed is that a showing by the defendant in a misinformation case that the plaintiff disbelieved the information will exonerate the defendant of liability. Admittedly, the rule does not work as well when there is a significant delay between the defendant's erroneous statement and the plaintiff's search for confirmation. 62

Another way in which the plaintiff's conduct may act as a bar to his recovery is through placing himself voluntarily - or more precisely, gratuitously - into a situation where there is heightened risk of emotional stress. This occurred in Justus v. Atchison, 63 a California bystander case in which the plaintiff's lost because they failed to meet precisely one of the Dillon guidelines. Plaintiff was an expectant father present in the delivery room, when through medical malpractice, the fetus was stillborn. We are concerned here solely with the claim of the plaintiff-father for negligently induced emotional harm, rather than the wrongful death claim.

Rigidly following Dillon, the plaintiff 64 is denied recovery because while he was a contemporaneous witness to some events in the delivery room, he did not observe any injury or the stillbirth itself. Plaintiff was unaware of any serious problem until the attending physician acknowledged the stillbirth outside of the delivery room. Thus,

60 Cases for erroneous notification of death seem to be hybrids of the ancient exceptions to the ban on recovery for emotional harm negligently inflicted - namely, corpse mishandling cases and faulty telegram cases. The corpse cases were always considered to give rise to unquestionably genuine harm, by their very nature. Erroneous wire cases were somehow related, but of course differ in that all that happened was an erroneous communication, not a mishandling. The common thread was death, the subject matter of the most common telegram cases.

61 There were two patients named Emma Johnson, and the "other one" died.

62 If the defendantcorrects before the plaintiff seeks confirmation, perhaps that should bar recovery also. The question of the defendant's correction is more general than consideration of attempted confirmation by the plaintiff. See supra, note 26.


64 There are a number of plaintiffs here because several cases with similar fact patterns were consolidated on appeal.
there was no shock induced by contemporaneous observation of a sudden event, and recovery was barred following *Dillon*.65

But for our purposes, what appears as dicta in *Justus*66 points out that the expectant father was a special kind of bystander. In *Justus*, the plaintiff had gratuitously entered the delivery room for no useful purpose other than the amorphous benefits of accompanying mother and witnessing the birth; that is, there is no cognizable benefit accruing to the newborn running from the father’s presence. Further, it is well known that emotions and potential stresses run high in such a place.67

These factors differ from the more typical situation, where the bystander is probably present in a protective capacity and not voluntarily there.68 Further, the environment may be termed "emotionally neutral", at least before the traumatic event, which is likely to be a collision. That is, the typical bystander case does not take place in surroundings where high emotion is commonplace.

The general rule ought to be that a gratuitous presence bars recovery for subsequent emotional harm; however, what I have called a gratuitous presence might be distinguished from a voluntary one. The two are similar in that both are positive elections of conduct, usually without specific reward to the actor, but they differ in that the voluntary act is intended to have a direct and beneficial effect on the potential victim. The distinction is clear if one contrasts a rescue attempt (clearly voluntary) to accompanying the expectant mother to delivery (gratuitous).

Thus, in summary, the gratuitous bystander should be barred as a matter of law from recovery, while the involuntary or voluntary (in the sense of a volunteer/rescuer) "bystander" may proceed.69

**FAILURE OF PLAINTIFF TO MITIGATE HARM**

One of the cases cited as a landmark in the establishment of emotional harm as an independent tort is *Bass v. Nooney Co.*, a 1983 case.70 The court decided for the first

---

65 The very natural and foreseeable grief at the loss of a child, by itself, is not compensable.

66 *Id.* at 578, 565 P.2d at 136, 139 Cal. Rptr. at 111. It is surely less common for a direct victim of misinformation to intrude his presence than for a bystander, but where he does, he should not recover emotional harm damages.

67 "Surely a layman who voluntarily observes a surgical operation must be prepared for the possibility of unpleasant or even harrowing experiences. This is no less true of...childbirth. The ever-present possibility of emotional distress dissuades us from extending the *Dillon* rule into the operating amphitheater in these circumstances." *Id.*

68 "By its nature the *Dillon* cause of action presupposes that the plaintiff was an involuntary witness to the accident." *Id.*

69 The term "bystander" is hardly appropriate for one who is in fact a rescuer; of course, in that context, the bystander label means he has no immediate relationship to the defendant.

70 646 S.W.2d 765 (Mo. banc 1983).
time in Missouri that a cause of action may be stated in the absence of physical impact. Let us examine the fact pattern of this case to see what disposition could be made, once we accept that impact is not a precondition to recovery. It must be said in preface that the failure of plaintiff to mitigate harm is probably not the most important aspect of this case. However, it is a factor, and discussion of the other factors may well be a useful review.

Mrs. Bass got stuck for 30 minutes in an elevator in the building where she was employed.\footnote{There is some uncertainty concerning the time Mrs. Bass actually was stuck in the elevator; the text of the opinion clearly counts the time as a half-hour. However, note 2 of the opinion indicates that the plaintiff estimated she had entered the elevator at an earlier time than other reports, and she may have thought it was an hour. I have used the half-hour in my discussion. \textit{Id.} at 766-7.} The defendant's negligence is shown via \textit{res ipsa loquitur}.\footnote{The acceptance of plaintiff's claim of \textit{res ipsa loquitur} by the court is rather facile; their discussion rests on a determination that "the jury could reasonably find that either or both of the defendants were in control of the elevator, so as to make the application of \textit{res ipsa loquitur} proper." (The defendants are the building owners and the Otis Elevator Company.) \textit{Id.} at 768. This is a rather strained interpretation; a minority opinion in this case points out that "in our complex modern society ever-increasing numbers of people are served daily by electrically energized devices and machines serving human needs..., devices that..., inevitably breakdowns...and failures of electrical and mechanical devices occur." \textit{Id.} at 777. There is a difference between the opinions, and the latter is the better view, in my opinion.} The issues we address are susceptibility, plaintiff conduct and seriousness.

Testimony was brought out that Mrs. Bass was under stress in her personal life, being in the midst of a divorce, and being responsible for three teenage children. That however is not the major point; the susceptibility is virtually self-proving. Here is a woman stuck in an elevator for thirty minutes who admits she did not fear physical injury by a fall of the elevator. She is merely stuck. Soon after she gets stuck, she summons help, and while she is not immediately freed, she is in communication with persons who call maintenance people. Would a normal person have been emotionally harmed after such a minor incident?\footnote{Note also that unlike many of the physical effects running through cases discussed herein, here there is no impact and it could be argued that the harm was temporary, if we correlate the extent of the harm to the extent of Mrs. Bass's involuntary confinement. This may be a way of saying that she is entitled to 30 minutes worth of emotional harm. Such an equation would not be appropriate where the harm was a truly serious one, such as a bystander observing the death of a relative, but seems appropriate here.} Perhaps this is a jury question, but there is more.

While the elevator did get stuck, the elevator did not malfunction in any other way. The alarm bell worked, as did the lights, and there was a "Door Open" button which plaintiff did not push. As pointed out in the dissenting opinion, plaintiff was aware of the function of this device.\footnote{646 S.W.2d 765, 780 (Mo. banc. 1983).} All plaintiff had to do was to push this button, and she would have been able to step out of the elevator easily, because it stopped only...
one foot above the nearest floor. Thus, the plaintiff did nothing to mitigate the distress; her inaction may be seen as a contributing factor. Plaintiff’s argument was that her fright was such that it rendered her incapable of such action. This argument stretches foreseeability too far. One must foresee that the victim is not so terrified that she is capable of ringing the alarm, but she is so terrified that she cannot push the "Door Open" button.

Several other facts are in the defendant’s favor. Elevators are very dangerous; if a cable breaks, or someone falls down the shaft through an open door, death is likely. Among what might be classed as secondary dangers is the potential for injury in extricating passengers from a stuck elevator. This is secondary in that the magnitude of the bodily injury is very likely to be minor, compared to what would happen in a fall down the shaft. Because elevators are dangerous they have numerous safety features and undergo routine inspections by law. No doubt, elevators could be made even safer, but only at prohibitive inconvenience or cost. Would not have Mrs. Bass’s injuries been prevented or ameliorated had the defendant provided an attendant operator or for that matter, a trained psychologist who would recognize and treat claustrophobia or acrophobia? The point is not to suggest these measures, but to emphasize that we must accept some risk. Because of the regulation of elevators by public authorities, there has probably been a weighing of what devices among the myriad possibilities should be installed.

In this case, all of the safety features worked as planned, but Mrs. Bass did not use them all; it seems very unjust to penalize the defendant where he has expressly provided devices to mitigate potential harm, and the plaintiff does not use them.
THE CASE OF "INTENT":  
SHOULD THE ELEMENTS OF MURDER BE EXPANDED IN VIRGINIA?  
David L. Thomas

INTRODUCTION

The word "murderer" connotes in human beings an intensity of feeling that ranks near the top of the human emotional spectrum. Although the "murderer" is to be abhorred, his status is to be protected. That is, the term "murderer" embraces such magnitude of horror that the label must be saved only for those who truly deserve it. Therefore, the state legislature has established strict rules concerning the designation "first degree murderer." Among the rules is one that demands that the slayer possess not only the express "intent to kill," but also a "willful, premeditated, and deliberate" mindset. My inquiry here examines whether that rule of first degree murder in Virginia can be replaced with either an "intention to cause serious bodily injury" or a "reckless disregard for human life."

DISCUSSION

HISTORY

As common law murder evolved, a debate raged, and remains, as to what constitutes first degree and second degree murder. 1 Each state has instituted its own

___________________
1 See generally R. PERKINS, PERKINS ON CRIMINAL LAW 86-96 (1969). Perkins discusses the origins of the degrees of murder, which were embodied in a 1794 Pennsylvania statute that came to be known as the "Pennsylvania Pattern." Today this doctrine is no longer embodied in an existing authoritative statute but survives in modern case law in Commonwealth v. Drum, 58 Pa. 9 (1868).

ATTEMPTED SECOND DEGREE MURDER--Does It Exist?

Generally when one speaks of attempted murder, the list of essential elements includes only those used in first degree murder offenses. Additionally, it is essential that in order to have a valid attempt charge that the crime was not accomplished. Thus it would seem that a charge of attempted second degree murder would be rendered an impossibility. (The malice in second degree murder is generally imputed from the unlawful killing.) However, this assumption may be too hasty.

The presumption may be that all unlawful killings are second degree murder, but the Commonwealth still has the burden of proving malice if the defendant shifts the burden back to the Commonwealth. Thus, in second degree murder cases the prosecution must either elevate the killing to first degree murder through a showing of willful intent to kill or it must battle the defendant's contention that there was no malice to constitute second degree murder by showing that the defendant's actions were either with reckless indifference to the value of human life, Commonwealth v. Malone, 354 Pa. 180, 47 A.2d 445 (1956), or that the defendant intended serious bodily injury which was "life threatening." Cruce v. State, 87 Fla. 406, 100 So. 264 (1924). Thus malice, it would seem, can be proven in second degree murder cases, even without a death. Therefore, the elements of attempted second degree murder would be:

1. Either an intent by the defendant to cause serious bodily harm that is life-threatening or a reckless indifference to human life demonstrated by the defendant's actions. (No "intent to kill" element is necessary in second degree murder cases in Virginia.)
views concerning the requirements for each degree of murder. For example, in most states a presumption of second degree murder arises from the death of an individual when caused by a criminal agent. The unlawful act causing the death imputes the necessary malice. However, this view of murder is not accepted by all states. Another difference is in the "intent" requirement. Most states require an "intention to kill" only for first degree murder; however, a minority of states require a specific "intent to kill" for second degree murder as well.

Not surprisingly, conflicts exist over the restrictiveness of the first degree murder statutes. Many jurists believe that the "intent to kill" and the "premeditated and deliberate killing" requirements are too narrow. These jurists are of the opinion that the intent or malice requirement of first degree murder can be fulfilled by an intention.

2. Actions that would set into motion the process of events leading up to the fulfillment of the crime.

3. No one need be injured by the attempt. (There is no corpus delicti element.) Because malice can be fulfilled for second degree murder without a killing, the door is left open for a valid charge and conviction of attempted second degree murder. However, the difficulty in using this theory lies in the use of recklessness; for by definition recklessness lacks the specific intent which is required by all crimes of attempt. That leaves attempted second degree murder limited to only those defendants exercising an "intent to cause serious injury".

It should be noted that the theory of attempted second degree murder has never been used in the state of Virginia.

An Example of Attempted Second Degree Murder:

A man is walking down the street. He carries a gun on his person for protection. Another man coming from the opposite direction bumps into the armed man and calls him a name. The armed man, who is provoked, but not to a level justifying voluntary manslaughter, draws his weapon and fires at the other man. The intent of the armed man is to seriously injure the other by shooting him in the chest. There is no specific intent to kill. However, the recklessness of the act itself raises a presumption of second degree murder malice. The shot fired at the passerby misses him and lodges in a tree. Because there is both malice and a specific intent to injure, the charge of attempted second degree murder is warranted.

2 R. PERKINS, supra note 1, at 88-9.


5 State v. Rowley, 216 Iowa 140, 248 N.W. 340 (1933). The court held that the State need prove more than a death by an unlawful act to raise a presumption of murder.

6 New York does not even require an "intent to kill" as an essential element of murder. See People v. Jernatowski, 238 N.Y. 188, 144 N.E. 497 (1924).

on the part of the accused to do "great bodily harm" to the victim or by a "reckless
disregard for human life" demonstrated by the accused. The influence these jurists
have on courts' interpretations of the differing state murder statutes is great. Today
the New York and federal courts, two of the largest systems in the country, lead the
way in expanding the intent requirement of first degree murder to include "reckless
disregard for human life," while Louisiana and Indiana have led the charge toward
the inclusion of the intent to do "serious bodily injury."

8 An example of this judicial thought may be found in People v. Murphy, 1 Cal.2d
37, 32 P.2d 635 (1934), where the court extended the torture element of first degree
murder to wife beating.

9 Jernatowski, 238 N.Y. at 188, 144 N.E. at 497.


The court in Shaw held that "the malice required for conviction of first degree or
second degree murder does not require [a] subjective intent to kill, but may be
established by evidence of conduct which is reckless and wanton and gross deviation
from [a] reasonable standard of care, of such nature that [the] jury is warranted in
inferring that [the] defendant was aware of [the] serious risk of death or serious bodily
harm." Shaw, 701 F.2d at 392 (quoting United States v. Black Elk, 597 F.2d 49, 51 (8th
Cir. 1978)).

Although the federal courts have stated that there still must be premeditation, that
requirement is presumed fulfilled from successfully showing that the defendant's act
constituted gross recklessness. In Shaw, the defendant negligently shot at a car while
hunting deer at night (spotlighting). Although there was no specific intent to kill or
even premeditation proven, the defendant was convicted of first degree murder on the
grounds that his act was grossly reckless. Shaw, 701 F.2d at 367.

See also Deputy v. State, 500 A.2d 581, 596 (Del. 1985), where the court held in regard
to felony-murder that "a person possessing a reckless state of mind can only be
convicted of first degree murder if he recklessly kills while committing a felony." Similar
to this, the court in State v. Brooks, 499 So.2d 741 (La. App. 1986).

11 In Brooks the defendant contendad that a specific intent to do serious bodily injury
could not replace the intent to kill element of first degree murder. The court disagreed
holding that "in proving first or second degree murder, either the specific intent to kill
or the specific intent to inflict great bodily harm can be proven". Brooks, 499 So.2d at
744. However, the court did sustain the defendant's argument on the grounds that
attempted first degree murder, of which the defendant was charged, required an intent
to kill, not an intent to do harm.

See also State v. Martin, 213 N.J.Super. 414, 517 A.2d 513 (1986). In Martin, where
the defendant killed a woman by knowingly setting a building on fire when he knew
people were inside of it, the court held that "a murder conviction based on "purposeful"
or "knowing" conduct can result from conduct which is practically certain to cause
serious bodily injury when death is a result of the injury caused". Id. at 418, 517 A.2d
at 517.

See also Robinson v. State, 453 N.E.2d 280 (Ind. 1983). In Robinson, the defendant
kicked his child to death for bed wetting. The court in upholding the first degree
murder conviction held that "there must be evidence that the defendant had a conscious
objective to kill the victim or was aware that his conduct would result in [the] death of
[the] victim". Id. at 280 (quoting Burkhalter v. State, 397 N.E.2d 596 (Ind. 1979)). An
Although Michie's Jurisprudence attempts to read the intent to do "serious bodily injury" into Va. Code Ann. 18.2-32, it does so in an unacceptable manner. Michie's contention is that the statute is satisfied because an intent to do "serious bodily injury" is equivalent to "lying in wait" for the victim, thus implying the intent-to-kill element. To support its deviation from the wording of the statute, Michie erroneously cites Commonwealth v. Jones, which does not even address the issue of "serious bodily injury." In fact, Jones is a landmark "lying in wait" case, and in it, the court states explicitly that an "intent-to-kill" element is still necessary in order to prove first degree murder and cannot be overlooked simply by proving that the accused was "lying in wait." The party must necessarily be "lying in wait" to kill. Therefore, the popular intent to kill is not necessary, only a presumed awareness by the reasonable man standard that the intent to harm would in fact occasion death. See Honesty v. Commonwealth, 81 Va. 283 (1886) for similar language.

For a listing of other cases affirming the trend toward inclusion of the "intent to do serious bodily injury" as a replacement for the "intent to kill" elements of first degree murder, see also State v. Gerrel, 499 So.2d 381 (La. App. 1986); State v. Flowers, 509 So.2d 588 (La. App. 1987); Burse v. State, 515 N.E.2d 1383 (Ind. 1987); Ortiz v. State, 651 S.W.2d 764 (Tex. Cr. App. 1983); and Smith v. State, 486 N.E.2d 465 (Ind. 1985).

The full text of the statute reads:

First and second degree murder defined; punishment.--

Murder, other than capital murder, by poison, lying in wait, imprisonment, starving, or by any willful, deliberate, and premeditated killing, or in the commission of, or attempt to commit, arson, rape, forcible sodomy, inanimate object sexual penetration, robbery, burglary or abduction, except as provided in Section 18.2-31, is murder of the first degree, punishable as a Class 2 felony.

All murder other than capital murder and murder in the first degree is murder of the second degree and is punishable as a Class 3 felony.


In Jones, the defendant and the deceased were socializing at a grocery in Lynchburg, Virginia in the spring of 1829. A prostitute, whom each man had known, entered the establishment and began insulting the deceased. The deceased thereupon struck the prostitute and she appealed to the sympathies of the defendant to protect her. This incident incited a two-day war of words and deeds between the two men which culminated in the defendant's purchasing a gun and shooting the deceased. The rule of law that this case demonstrates is "intent". The court, in upholding the defendant's conviction of first degree murder, continually showed how the defendant had planned and deliberated over the killing. Id. at 598.

Id.
argument imputing intent to do "serious bodily injury" to those "lying in wait" is very weak.

However, Michie's conclusion is invalid only in its analysis. The conclusion is correct! The intent to do "serious bodily injury" does replace the "intent to kill" and "premeditated and deliberate killing" requirements in first degree murder in Virginia. The reasoning of the Virginia Supreme Court is not that it is incorporated into the "lying in wait" element, but that it raises a presumption of deliberation. The court limits the application of the intent to do "serious bodily injury" to circumstances where a reasonable man would have known that the harm intended would "probably occasion the victim's death." This expansion of first degree murder was first instituted in *Honesty v. Commonwealth* and was reaffirmed in *Hall v. Commonwealth*. The full rule of law reads: 

> "[a]nd if there be a reasonable doubt whether he had willed, deliberated, and premeditated to kill the deceased, or to do him some serious bodily injury, which would probably occasion his death, the jury ought not to find him guilty of murder in the first degree." 

---

16 *Honesty v. Commonwealth, 81 Va. 283 (1886).*

17 *Id. at 283.*

A complete narrative of *Honesty* follows:

On November 14, 1884 there was a celebration in an unspecified Virginia city. The defendant was out in the street with an associate "halloowing and dancing". *Honesty, 81 Va. at 302.* The defendant had in his possession a large stick and threatened aloud that "if any son of a bitch of a democrat as much as rubs against me, I will give him hell." *Id. at 302.* Later that same evening the deceased and a friend proceeded toward the friend's home. They stopped along the way in order to enter an alley to "answer a call of nature". *Id. at 303.* There, they heard the defendant and his compatriot arguing. When the deceased requested the defendant to end the heated argument, the defendant began harassing, cursing, and pushing the deceased. When the deceased attempted to leave the alleyway, the defendant struck him in the head with a brick, thereby causing his immediate death.

The court held that the defendant's conviction of first degree murder should be sustained on the grounds that the use of the brick as a deadly weapon raised the presumption that the defendant deliberately caused the deceased's death. This the court reasoned was because a reasonable man would know that a blow, such as the one the defendant gave the deceased, would "probably occasion death". *Id. at 294.*

18 *Hall v. Commonwealth, 89 Va. 171, 15 S.E. 517 (1892).*

In following *Honesty*, the court in *Hall* stated that a serious wound caused by a deadly weapon raised a *prima facie* presumption of an intent to do "serious bodily injury" and would substitute for the "willful, premeditated, and deliberate" element of first degree murder. *Hall, 89 Va. at 178, 15 S.E. at 517.* This finding was affirmed by the court in *Mealy v. Commonwealth, 135 Va. 585, 115 S.E. 563 (1923).* See also *Wade v. Commonwealth, 202 Va. 177, 116 S.E.2d 99 (1960)*, where the court refers to proof of intent to do "serious bodily injury" as an alternative to the "willful, premeditated, and deliberate" element of first degree murder.

19 *Honesty, 81 Va. at 294 (emphasis added).*
The case law in Virginia is not as clear when the charge is first degree murder as the result of "reckless behavior." No major cases in Virginia address reckless

A Logical Expansion of the Law?

The logic behind the expansion of the first degree murder elements to include the intent to do "serious bodily injury" is based in part upon the court's view of premeditation as seen in Honesty. The array of cases that cite Honesty as controlling refer to the case's alternative interpretation of premeditated and deliberate murder. See Hall v. Commonwealth, 89 Va. 171, 15 S.E. 517 (1892), Mealy v. Commonwealth, 135 Va. 585, 115 S.E. 563 (1923), and Carson v. Commonwealth, 188 Va. 398, 49 S.E.2d 704 (1948).

In Hall, the use of a firearm in and of itself was sufficient to justify Honesty's alternative murder element. Hall, 89 Va. at 178, 15 S.E. at 519. The logic behind the expansion is embodied in the latter part of Honesty's definition of the crime of first degree murder, which may be summarized as: "If a reasonable man knows his actions could result in death, than he should be held responsible for the natural consequences of those acts."

The argument for the expansion of first degree murder is in fact the same argument used by the Michigan Supreme Court in People v. Aaron, 409 Mich. 672, 299 N.W.2d 304 (1980), where the court struck down the felony-murder doctrine. The argument is that the law must punish not in accordance with mere presence, but with culpability. One must be made responsible for culpable intent and the actions that follow from that intent. Although a defendant can always attempt to reduce first degree murder to second degree murder by stating that he "didn't mean to kill," the state must logically be entitled to a presumption of intent if a reasonable man would know that his actions would probably occasion the death of the other party. One may not be allowed to ignore the consequences of ones act by mere denial after the fact. Intent is to be proven by the totality of the circumstances, Beck v. Commonwealth, 2 Va. App. 170, 342 S.E.2d 642 (1986), not by the per se testimony of the accused. To allow such an erroneous standard would make criminal intent obsolete.

How Negligence Fits into the Spectrum of Legal Liability

If the proposition that gross recklessness can give rise to first degree murder is true, than it is also a truism that negligence (the failure to exercise due care) is a wrong whose spectrum extends from the civil court proceeding to the highest realms of the criminal court system. A diagramming of that negligence spectrum follows:

<table>
<thead>
<tr>
<th>Civil Court Proceeding</th>
<th>/</th>
<th>Criminal Court Proceeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Negligence</td>
<td>Involuntary Manslaughter</td>
<td>Second Degree Murder</td>
</tr>
<tr>
<td>Negligence</td>
<td></td>
<td>First Degree Murder</td>
</tr>
</tbody>
</table>

Absence of Malice Malicious Intent
behavior as a replacement for the "intent to kill" element of first degree murder. Even Michie's is silent on the subject. Instead, persuasive law such as the landmark New York case People v. Jernatowski, which states that acts that show "reckless disregard for human life" can fulfill the requirement of intent for first degree murder, must be cited as authority in order to make a case for this doctrine in the Virginia courts.

The viability of using reckless conduct as a substitute for the "intent to kill" and the modus operandi (premeditation) in a first degree murder case is not as unusual as it would first appear. The reasoning of a court effectuating the substitution may be grounded in common sense. If the actions of the defendant are truly outrageous enough to elevate the crime to first degree murder, than perhaps the "intent" and motive do, in fact, exist and, hence, may be implied for purposes of proving murder.

It would seem logical that at some point on the negligence-recklessness spectrum, a reasonable man knows he is committing a dangerous act and understands its foreseeable consequences. At such a level the defendant, whether he had actually contemplated his intention, would be implicitly intending the outcome of his actions. The only true defense a defendant could raise, after the Commonwealth had elevated the recklessness to the first degree level and, thereby satisfied its burden of proof, would be the irresistible impulse insanity defense. (That is the affirmative defense where although the defendant knows of the consequences of his actions, he is nevertheless unable to stop his behavior. An impulse compels him to act against his will.)

Allowing an individual's outrageous acts either to slide by the judicial scales of criminal prosecution or to avoid higher levels of criminal punishment by pleading a lack of "intent," in all likelihood gives rise to intelligent killers with outrageous and uniquely planned premeditated murders.

However, the scenario just contemplated can be avoided if the Commonwealth is allowed to raise a presumption of "intent" by proving a stipulated level of outrageousness. The burden is then shifted onto each defendant to prove a lack of "criminal intent." This presumption, according to some jurists, would lead to a fairer system of adjudicating crimes involving death.

On the other hand, the difficulty with this type of system lies in the fact that it raises first degree murder to a strict liability or quasi-res ipsa loquitur standard. Those types of classes of proof in a civil system may be acceptable, but under the guise of criminal justice they are unconstitutional in their disregard for due process.

Therefore, the premise onto which extreme gross recklessness as a substitute for first degree murder may be based is fatally flawed. Unless due process can be satisfied, reckless conduct in and of itself can never raise a presumption of first degree murder no matter what level of recklessness is proved by the Commonwealth. In the opinion of this author, the New York Murder Statutes as they are presently interpreted by New York Courts through decisions like Jernatowski are constitutionally unsound.

21 The Logic Behind Recklessness
(In its context as a substitute for first degree intent)

The viability of using reckless conduct as a substitute for the "intent to kill" and the modus operandi (premeditation) in a first degree murder case is not as unusual as it would first appear. The reasoning of a court effectuating the substitution may be grounded in common sense. If the actions of the defendant are truly outrageous enough to elevate the crime to first degree murder, than perhaps the "intent" and motive do, in fact, exist and, hence, may be implied for purposes of proving murder.

It would seem logical that at some point on the negligence-recklessness spectrum, a reasonable man knows he is committing a dangerous act and understands its foreseeable consequences. At such a level the defendant, whether he had actually contemplated his intention, would be implicitly intending the outcome of his actions. The only true defense a defendant could raise, after the Commonwealth had elevated the recklessness to the first degree level and, thereby satisfied its burden of proof, would be the irresistible impulse insanity defense. (That is the affirmative defense where although the defendant knows of the consequences of his actions, he is nevertheless unable to stop his behavior. An impulse compels him to act against his will.)

Allowing an individual's outrageous acts either to slide by the judicial scales of criminal prosecution or to avoid higher levels of criminal punishment by pleading a lack of "intent," in all likelihood gives rise to intelligent killers with outrageous and uniquely planned premeditated murders.

However, the scenario just contemplated can be avoided if the Commonwealth is allowed to raise a presumption of "intent" by proving a stipulated level of outrageousness. The burden is then shifted onto each defendant to prove a lack of "criminal intent." This presumption, according to some jurists, would lead to a fairer system of adjudicating crimes involving death.

On the other hand, the difficulty with this type of system lies in the fact that it raises first degree murder to a strict liability or quasi-res ipsa loquitur standard. Those types of classes of proof in a civil system may be acceptable, but under the guise of criminal justice they are unconstitutional in their disregard for due process.

Therefore, the premise onto which extreme gross recklessness as a substitute for first degree murder may be based is fatally flawed. Unless due process can be satisfied, reckless conduct in and of itself can never raise a presumption of first degree murder no matter what level of recklessness is proved by the Commonwealth. In the opinion of this author, the New York Murder Statutes as they are presently interpreted by New York Courts through decisions like Jernatowski are constitutionally unsound.

22 238 N.Y. 188, 144 N.E. 497 (1924). In Jernatowski, a railroad strike occurred in Buffalo, New York. The deceased's husband was a foreman with the railroad. He did not participate in the strike and elected to cross the picket line. Late one evening one of the strikers fired two gunshots into the foreman's house in an attempt to scare the foreman. The defendant at the time of the shooting knew that the house was occupied and had even been told moments before the incident by the deceased to "get away from there." The result of the defendant's act was the death of the foreman's wife, as one of the gun shots hit and killed her instantly.

The court held that the defendant's act justified a conviction for first degree murder because the defendant did "an act imminently dangerous to others, and evincing a depraved mind, regardless of human life, although without a premeditated design to effect the death of any individual." Id. at 190, 144 N.E. at 497.
CONCLUSION

Thus "if A finds B asleep on straw, and lights the straw, meaning to do B serious bodily injury, but not to kill him, and B is burned to death, it is murder" (in the context of first degree murder).\textsuperscript{23} The intent to do "serious bodily injury" replaces the "intent to kill" element under Virginia's first degree murder statute. However, the argument over "reckless disregard for human life" is still an open question in Virginia.

\textsuperscript{23} Honesty, 81 Va. at 295.
APPENDIX:

PROPOSED JURY INSTRUCTIONS FOR FIRST DEGREE MURDER IN VIRGINIA

It is my contention that the Virginia Model Jury Instructions should be amended to reflect an intent to do "serious bodily injury" as a viable substitute element for the crime of first degree murder. The corrected version should read:

The defendant is charged with the crime of first degree murder. The Commonwealth must prove beyond a reasonable doubt each of the following elements of the crime:

(1) That the defendant killed (Name of Person); and
(2) That the killing was malicious; and
(3) That the killing was willful, deliberate, and premeditated or that the intention of the defendant was to cause (Name of Person) serious bodily injury, which would probably occasion his death.\(^{24}\)


The unedited current version of the Virginia Model Jury Instructions includes all of the proposed new instructions minus the wording "or that the intention of the defendant was to cause (Name of Person) serious bodily injury, which would probably occasion his death." The new wording comes from the jury instructions given by the court in Honesty, 81 Va. at 294.