VOLUME 1992 NUMBER 3

UTAH LAW REVIEW

SYMPOSIUM: ETHICS, BIOETHICS, AND FAMILY LAW

UNIVERSITY OF UTAH COLLEGE OF LAW

UTAH LAW REVIEW

VOLUME 1992 NUMBER 3

EDITORIAL OFFICE: The *Utah Law Review* is published at the University of Utah College of Law, 325 University Street, Salt Lake City, Utah 84112 by the Utah Law Review Society. Second-class postage paid at Salt Lake City, Utah, and at additional mailing offices. All communications should be sent to the editorial offices. Postmaster: Please send form 3579 (change of address) to *Utah Law Review*, University of Utah College of Law, Salt Lake City, Utah 84112.
SUBSCRIPTIONS: The *Utah Law Review* is published quarterly. The subscription rate is $20.00 for one year. Single issues may be obtained for $6.00 per issue and reprints of articles and notes for $2.50 per copy. If the subscriber wishes a subscription discontinued at its expiration, notice should be sent; otherwise, it is assumed that continuation is desired.
BACK ISSUES & REPRINTS: Complete volumes and individual issues of the *Utah Law Review* before the current year are available from William S. Hein & Co., Inc., 1285 South Main Street, Buffalo, New York 14201. To order by telephone, call 1-800-828-7571. If in New York, call (716) 882-2600.
 Printed August 1992 (ISSN 0042-1448).

Printed in the United States of America
by Darby Printing Company

UNIVERSITY OF UTAH
COLLEGE OF LAW FACULTY & ADMINISTRATION

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

AGUILAR, JR., REYES (1992), Director of Admissions and Financial Aid. B.A., 1988, Texas A&M University; J.D., 1992, University of Utah.

BENNETT, WALLACE R. (1953-1986), Professor Emeritus of Law. B.A., 1948, M.S., 1950, J.D., 1953, University of Utah; LL.M., 1957, Harvard University; LL.M. (Trade Regulations), 1963, New York University.

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

CASSELL, PAUL G. (1991), Associate Professor of Law. B.A., 1981, J.D., 1984, Stanford University.

DICKEY, JEANNE HENDERSON (1992), Assistant Dean of the College of Law. B.A., 1970, University of Minnesota; J.D., 1977, University of Utah.

DYER, BOYD KIMBALL (1971), Professor of Law. B.A., 1962, Stanford University; LL.B., 1968, Harvard University.

EMERY, ALFRED C. (1947), Distinguished Professor Emeritus of Law and President Emeritus of the University of Utah. B.S., 1940, J.D., 1947, University of Utah.

ENGLE, KAREN L. (1992), Associate Professor of Law. B.A., 1984, Baylor University; J.D., 1989, Harvard University.

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

FLORES, ROBERT L. (1990), Associate Professor of Law. B.A., 1978, J.D., 1987, University of Utah.

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

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

FRANCIS, LESLIE PICKERING (1982), Professor of Law and Associate Professor of Philosophy. B.A., 1967, Wellesley College; Ph.D., 1974, University of Michigan; J.D., 1981, University of Utah.

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

GREENWOOD, DANIEL J.H. (1991), Associate Professor of Law. A.B., 1979, Harvard University; J.D., 1984, Yale University.

KOGAN, TERRY S. (1984), Professor of Law. B.A., 1971, Columbia College; B. Phil., 1973, Oxford University; J.D., 1976, Yale University.

LAHEY, KATE (1988), Director, Legal Writing Program and Adjunct Professor of Law. B.A., 1975, J.D., 1979, University of Utah.

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

LUND, THOMAS (1978), Professor of Law. B.A., 1964, Harvard University; LL.B., 1967, Columbia University; Ph.D., 1978, Oxford University.

MARTINEZ, JOHN (1984), Professor of Law. B.A., 1973, Occidental College; J.D., 1976, Columbia University.

MATHESON, JR., SCOTT M. (1985), Associate Dean for Academic Affairs of the College of Law and Professor of Law. B.A., 1975, Stanford University; M.A., Oxford University; J.D., 1980, Yale University.

MCCORMACK, WAYNE (1978), Professor of Law. B.A., 1966, Stanford University; J.D., 1969, University of Texas.

MITCHELL, BONNIE L. (1987), Associate Dean of the College of Law. B.A., 1981, J.D., 1984, University of Utah.

MORRIS, JOHN K. (1979), Counsel to the President of the University of Utah and Professor of Law. B.A., 1966, University of California at Los Angeles; J.D., 1969, University of California at Berkeley.

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

POULTER, SUSAN R. (1990), Associate Professor of Law. B.S., 1965, Ph.D., 1969, University of California at Berkeley; J.D., 1983, University of Utah.

REUSCH, RITA T. (1986), Director, Law Library and Professor of Law. B.A., 1972, University of Minnesota; J.D., 1975, University of Idaho; M.L.L., 1978, University of Washington.

SCHMID, ROBERT L. (1954), Professor Emeritus of Law and former Director of the Center for Energy, Public Lands, and Environmental Law. J.D., 1954, University of Utah; M.L.L., 1956, University of Washington.

SKIBINE, ALEXANDER TALLCHIEF (1989), Associate Professor of Law. B.A. 1973, Tufts University; J.D., 1976, Northwestern University.

SMITH, LINDA F. (1984), Associate Professor of Law and Clinical Program Director. B.A., 1973, Ohio State University; J.D., 1976, Yale University.

SPURGEON, EDWARD D. (1980), William H. Leary Professor of Law and Policy, former Dean of the College of Law. A.B., 1961, Princeton University; LL.B., 1964, Stanford University; LL.M., 1968, New York University.

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

TEITELBAUM, LEE EDWARD (1986), Dean of the College of Law and Professor of Law. B.A., 1963, Harvard College; LL.B., 1966, Harvard University; LL.M., 1968, Northwestern University.

THREEDY, DEBORA L. (1986), Professor of Law. B.A., 1973, Beloit College; J.D., 1980, Loyola University of Chicago.

THURMAN, SAMUEL D. (1962-1986), Distinguished Professor Emeritus of Law, former Dean of the College of Law. A.B., 1935, University of Utah; J.D., 1939, Stanford University; L.L.D., 1984, McGeorge; L.L.D., 1988, University of Utah.

UTAH LAW REVIEW

VOLUME 1992 NUMBER 3

1991-92 BOARD OF EDITORS

****

UTAH LAW REVIEW

VOLUME 1992 NUMBER 3

TABLE OF CONTENTS

SYMPOSIUM: ETHICS, BIOETHICS, AND FAMILY LAW

COMMENTS

NOTES

Foreword

The Articles appearing in this issue are the written component of a symposium on Ethics, Bioethics, and Family Law held at the University of Utah College of Law on April 23 and 24, 1992. The participants included commentators on the family and bioethics from a variety of disciplines. The Symposium was attended by an equally diverse gathering of practitioners, professors, and students.

The Symposium was conducted in three panels. Each panel consisted of three participants and was followed by an open discussion of the issues presented. A second aspect of the Symposium included a public forum on ethics and family decision making, during which all of the participants joined in a discussion moderated by Lee E. Teitelbaum, Dean and Professor of Law at the University of Utah College of Law, and Carl E. Schneider, Professor of Law at the University of Michigan Law School. The Articles that follow are edited versions of those presented at the Symposium. They appear in the order they were presented.

The first panel discussed questions related to family responsibility. In his paper, *Bioethics and Fatherhood*, Dr. Daniel Callahan, co-founder and Director of the Hastings Center, discusses the diminishing role of fathers and fatherhood in our society. He argues that the role of the father must be "revived and reinvigorated" to avoid long-term harm to men, women, and children, both individually and in their relationships. In *Frail Parents, Robust Duties*, Hilde Lindemann Nelson, Associate Editor of the *Hastings Center Report* and Dr. James Lindemann Nelson, Associate for Ethical Studies at the Hastings Center, examine the sources and difficulties of filial obligation in the context of caring for aging parents. They contend that children have a duty to care for their parents, but reject commonly held views that this duty is based solely on indebtedness, consent, or gratitude. Instead, they argue that the duty stems from the intimacy that inheres in the parent-child relationship. Dean Teitelbaum also addresses filial obligations in *Intergenerational Responsibility and Family Obligation: On Sharing*. He reviews the difficulties of directly regulating family relations through law and criticizes family responsibility statutes that shift the burden of familial care from the public to the private realm. Instead, he asserts, the complex problems of family obligations are best met through shared public and private efforts.

The second panel discussed issues relating to the quality of family decision making. Dr. John Hardwig, Professor of Philosophy and Humanities and Professor of Medical Ethics at East Tennessee State University, discusses the problem of whose interests should properly be considered in proxy decision making in *The Problem of Proxies with Interests of Their Own: Toward a Better Theory of Proxy Decisions*. He states that, while current medical ethics formally focuses on the autonomy of the patient, proxy decision makers generally consider factors other than the patient's best interest. He also argues that consideration of factors other than the wishes of a patient is morally appropriate. In *Bioethics and the Family: The Cautionary View from Family Law*, Professor Schneider examines the extent to which family law can inform bioethical dilemmas through its experience in delineating the "ethical and social bases of family life." Although he concludes that the inherent limits of family law as a tool of regulation leave little to directly inform the field of bioethics, the sources of family law's failings can still teach bioethics a great deal about modern medical dilemmas. In an effort to clarify the standards guiding proxy decision makers, Patricia D. White, Visiting Professor of Law at the University of Michigan, presented her paper *Appointing a Proxy Under the Best of Circumstances*. Professor White examines the wide variety of choices available to competent individuals when appointing proxy decision makers. She argues that careful examination of available choices will not only assist people in more clearly expressing their desires to their proxy, but will also aid the medical decision making process when no proxy has been appointed.

The final panel considered medical decisions and family issues. In *The Roles of the Family in Making Health Care Decisions for Incompetent Patients*, Dr. Leslie P. Francis, Professor of Law and Associate Professor of Philosophy at the University of Utah, explores the role of the family in making health care decisions for incompetent family members. She concludes that many reasons weigh in favor of including the family in such decisions, but that no one model of family participation can adequately address all circumstances. Instead, family roles should differ, depending on whether the patient is a formerly competent adult, a never competent adult, or a child. Dr. Jay A. Jacobson, Professor of Medicine at the University of Utah School of Medicine and Chief of the Division of Medical Ethics at LDS Hospital in Salt Lake City, inquires how doctors consider a patient's family in

making medical decisions. His paper, *Medical Ethics and Family Affairs*, views the outcome of this question as governed, at least in part, by how doctors resolve the tension between their duties to a patient's family and their duty of confidentiality to the patient. He then discusses how formal training guides physicians with respect to this tension, how physicians generally behave, and what factors other than professional norms influence their behavior. In *Telling Medical Stories: Sharing Information Among Doctors, Patients, and Families*, Martha Minow, Professor of Law at Harvard Law School, approaches the ethics of family medical decision making through narratives illustrating how doctors, patients, and families communicate and interpret medical information. She identifies impediments to clear communication whose avoidance would assist doctors in conveying information in a way that individual patients and their families are able to understand.

The Symposium was the result of the combined efforts of many individuals. The *Utah Law Review* would like to thank all of the participants and attendees for their contributions. The *Review* would also like to thank the Albert and Elaine Borchard Foundation for its generous financial support.

The genesis of the Symposium, including its content and organization, lies with Dean Teitelbaum and Professor Schneider. The *Review* is grateful to both for their efforts. We are especially indebted to Professor Schneider. It is rare for faculty members to understand collegiality so broadly as to include a law school and a law journal outside their own institution. We are grateful for his time and for his commitment to encouraging discussion of these critical issues facing the family.

The University of Utah College of Law and the *Utah Law Review* will continue to provide a forum for such discussions and hope to make symposia on family law a biennial event. In pursuit of this goal, Professor Schneider and Dean Teitelbaum are currently planning a symposium reflecting on various aspects of a quarter-century of divorce reform that will be held in the fall of 1993.

1992-93 Board of Editors

Bioethics and Fatherhood

*Daniel Callahan**

I. INTRODUCTION

For most of the rest of our culture, the twin issues of the meaning of masculinity (or maleness, depending on your tastes), and the significance of fatherhood are well-developed topics of public discussion. Whether as a response to feminism, on the one hand, or to independent uncertainties about what it means to be a male, on the other, the question of masculinity attracts considerable attention. While fatherhood was not exactly a neglected topic in years past, there seems little doubt that the nasty phenomena of more and more single-parent families, mainly headed by females, and a growing number of absent and neglectful fathers, has given the issue a fresh urgency. What does it mean to be a father? What is the importance of the father for the nurturing of children? What can be done to encourage and assist more responsible fatherhood? What is the relationship between fatherhood and masculinity?

These are interesting and important questions, and timely as well. One would, however, never guess that from reading the literature of bioethics. For whatever reason, that literature, when it focuses on gender at all, is almost exclusively interested in women. And when it focuses on parenthood, it almost exclusively focuses on motherhood. While the general topics of reproductive choices and artificial means of reproduction have had a central place in bioethics, the literature and debate have usually centered on women's choices or women's role in such things as surrogate motherhood and *in vitro* fertilization. Fathers and fatherhood are just absent from the discussion altogether.

The absence of fatherhood in the debate is puzzling, especially since the topic of artificial means of reproduction is a central one in the field. My surmise is that, because those means of reproduction depend so heavily upon anonymous male sperm

* Director, The Hastings Center, Briarcliff Manor, New York.

donations, and since such donations are rarely questioned for their moral propriety, there has been no need or place to talk about fathers. They just don't really count in that brave new world of reproduction. I will return later to that topic. Of more general importance is whether fatherhood can be given a fresh look and a reinvigorated role in bioethics.

At the heart of the problem and future of parenthood, and thus of the most basic and indispensable kind of human nurturing, is a *relationship*, of men, women, and children bound together. Professionals seem to have lost a sense of and feel for that relationship—of the way men, women, and children need and best flourish in the company of the other. Instead, professionals have done conceptually what society has been doing legally and socially—treating men, women, and children as separate and distinguishable, with their own needs and rights. Thus we now speak easily of women's rights, and children's rights, and (hardly surprising, even if amusing) we have seen the growth of a men's rights movement. Doubtless there are some good reasons for this fragmenting development, the most important being the way earlier generations were prone to stack the family relationship, and its ground rules, too heavily in favor of men; or, where children were concerned, to treat them too much as the property of their parents, not as persons in their own rights.

But it is time for some reintegration. The fragmentation is, unless corrected in the long run, going to be harmful for men, women, and children, both individually and in their relationship. A revived and reinvigorated place for fathers and the institution of fatherhood is as good a place as any to begin. I want to develop three points: (1) biological fatherhood carries with it permanent and nondispensable duties; (2) the rapid and widespread acceptance of artificial insemination donors was much too thoughtless and casual, but for just that reason symbolic of the devaluation of fatherhood; and (3) feminism as a movement has hurt both men and children, but also women, by its tendency to substantively displace fathers from a central role in the making of procreation decisions.

II. The Duties of Fatherhood

I begin here with the most simple and primitive of moral axioms, rarely articulated as such but as undeniable as anything can possibly be in ethics. The axiom is this: Human beings bear a moral responsibility for those voluntary acts that have an impact on the lives of others; they are morally accountable for such acts. I will not discuss the many nuances and problems that this axiom raises: what counts as "voluntary," how great must be the impact upon others, and which effects of actions on others are morally more or less important.

In the case of biological fatherhood those nuances will not ordinarily be of great importance. From this moral axiom I will argue that given the obvious importance of procreation in bringing human life into existence, fathers have a significant moral responsibility for the children they voluntarily procreate. What human action could be more important than that which creates new life, the burden of which the newly born person must live with for the rest of his or her life? What causal connection could be more direct than biological procreation, without which human existence would not be possible? A father can hardly be held wholly responsible for *what* a child becomes—much will depend upon circumstances—but a father can be held responsible with the mother for the fact the child comes to be at all.

One philosopher has advanced the notion that our only serious moral obligations are those we voluntarily impose upon ourselves, as in specific contracts.[1] There cannot be, she says, involuntary obligations. This is not the place to debate the full implications of such a theory—which must systematically close its eyes to what it means to live in a community with other people—but it is pertinent to make a single point. Unless a male is utterly naive about the facts of procreation, to engage in voluntary sexual intercourse is to be responsible for what happens as a result. To enter into a contract with another is, at the least, to undertake a voluntary activity with a known likely outcome. Sexual intercourse for an informed male is fairly close to that, so even on a contract theory of moral obligation, intercourse shares many critical features with a contract. Society, curiously, seems to

1. Judith J. Thompson, *A Defense of Abortion*, 1 PHIL. & PUB. AFF. 47, 65 (1971).

have been faster in establishing the moral and causal links between drinking and driving than between sexual activity and pregnancy. But that may be because society prefers to think that accidental, unwanted pregnancies come more from contraceptive ignorance and failure than from the sexual activities that require them; the former is a more comforting thought to sustain the sexual revolution.

From my moral axiom, therefore, and from what we know about the biology of human procreation, I believe there is no serious way of denying the moral seriousness of biological fatherhood and the existence of moral duties that follow from it. The most important moral statement might be this: Once a father, always a father. Because the relationship is biological rather than contractual, the natural bond cannot be abrogated or put aside. I conclude, that just as society cannot put aside the biological bond, so neither ought it put aside the moral bond, the set of obligations that go with that biological bond. If there are to be moral duties at all, then the biological bond is as fundamental and unavoidable as any that can be imagined.[2] Does this mean that each and every father has a full set of moral obligations toward the children he procreates? My answer is yes—unless he is mentally or financially incompetent to discharge those duties. To treat the matter otherwise is to assume that fatherhood *is* some kind of contractual relationship, one that can be set aside by some choice on the part of the father, or the mother and father together, or on the part of the state. This position does not preclude allowing one person to adopt the child of another, to play the role of father with a legal sanction to do so. This arrangement, however, is legitimate only when there are serious obstacles standing in the way of the biological father playing that role himself. Even then, however, he remains the biological father, and should the alternative arrangements for the child fail, he is once again responsible, and responsible whether he likes it or not, accepts it or not. The obligation stems from his original, irreversible act of procreation; so too is his moral obligation irreversible.

Imagine the following scenario. A father has, through the assorted legal ways society allows fathers to turn over their

2. James L. Nelson, *Parental Obligations and the Ethics of Surrogacy: A Causal Perspective,* 5 PUB. AFF. Q. 49 (1991).

parental authority to another, legally ceased to act as a father and someone else is caring for the child. But imagine that the other person fails to adequately act as a father; fails, that is, to properly care for and nurture the child. The child then returns to the father and says: "You are still my father biologically; because of you I exist in this world. I need your help and you are obliged to give it to me." I have never been able to imagine even *one* moral reason why a father in that circumstance could disclaim responsibility, and disclaim it if, even in principle, there was someone else available who could take care of the child. A father is a father is a father.

III. Fatherhood and Artificial Insemination

I find it remarkable that, with hardly any public debate at all, the practice—indeed, institution—of artificial insemination by an anonymous male donor so easily slipped in. What could society have been thinking about? In this section I will argue that it is fundamentally wrong and should have no place in a civilized, much less a supposedly liberal society. It is wrong for just the reasons I have sketched in Section II about the moral obligations that go with fatherhood. A sperm donor whose sperm is successfully used to fertilize an ovum, which ovum proceeds through the usual phases of gestation, is a *father*. Nothing more, nothing less. He is as much a father biologically as the known sperm inseminator in a standard heterosexual relationship and sexual intercourse.

If he is thereby a biological father, he has all the duties of any other biological father. It is morally irrelevant that (1) the donor does not want to act as a father, (2) those who collect his sperm as medical brokers do not want him to act as a father, (3) the woman whose ovum he is fertilizing does not want him to act as a father, and (4) society is prepared to excuse him from the obligations of acting as a father. Fatherhood, because it is a biological condition, cannot be abrogated by personal desires or legal decisions. Nor can the moral obligations be abrogated either, unless there are reasons why they *cannot* be discharged, not simply that no one wants them to be discharged. Just as a "surrogate mother" is not a "surrogate" at all but a perfectly real and conventional biological mother, so also is a sperm donor whose

sperm results in a child a perfectly real and conventional biological father.

Why was it decided to set all that aside? Why was it deemed acceptable for males to become fathers by becoming sperm donors but then to relieve them *totally* of all responsibility of being fathers, leaving this new father ignorant of who his child is and the child ignorant of who the father is? I was not present at that great cultural moment, but two reasons seem to have been paramount.

First, it was introduced under medical auspices and given a medical legitimation. Artificial Insemination by a Donor ("AID"), one author wrote, is "medically indicated in instances of the man's sterility, possible hereditary disease, rhesus incompatibility, or in most cases of oligospermia."[3] "Medically indicated?" But it does not cure anyone's disease—not some other would-be father who is sterile, or the woman who receives the sperm who is perfectly capable of motherhood without donated sperm. What is cured, so to speak, is a couple's desire to have a child; but medicine does not ordinarily treat relational problems (save in psychotherapy), so there is no reason to call the matter medical at all. Moreover, of course, since artificial insemination only requires a single syringe, inserted in a well-known place, there is nothing "medical" even about the procedure.

As Daniel Wikler has nicely pointed out, the professional dominance of doctors in the history of AID is a perfect case of the medicalization of a nonmedical act, and the establishment of a medical monopoly and legitimization as a result.[4] Just how far this medicalization has gone can be seen by the very language used to describe the procedure: "[Artificial Insemination] is of two basic types: homologous, when the semen is obtained from the husband (AIH); and heterologous, when the semen is acquired from a donor (AID)."[5] I wonder how many males, working pleasurably to produce some sperm, understood themselves to be engaged in a heterologous activity? There is very little that medical science cannot dress up with a technical term.

---

3. Mark S. Frankel, *Reproductive Technologies: Artificial Insemination*, 4 ENCYCLOPEDIA OF BIOETHICS 1439, 1444 (Warren T. Reich ed., 1978).

4. Daniel Wikler & Norma J. Wikler, *Turkey-baster Babies: The Demedicalization of Artificial Insemination*, 69 MILBANK Q. 5, 8 (1991).

5. *See* Frankel, *supra* note 3, at 1444.

The second reason for ready acceptance was probably that, in the name of helping someone to have a child, society seems to be willing to set aside any existing moral restraints and conventions. Perhaps in an underpopulated world, whose very existence is threatened by low birth rates, a case for artificial procreation might be made.

But it is hard to see why, in our world, where the problem of feckless and irresponsible male procreators is far more of a social crisis, society lets that one pass. One can well understand the urge, often desperate, to have a child. But it is less easy to understand an acceptance of the systematic downgrading of fatherhood brought about by the introduction of anonymous sperm donors. Or perhaps it was the case that fatherhood had already sunken to such a low state, and male irresponsibility was already so accepted, that no one saw a problem. It is as if everyone argued: Look, males have always been fathering children anonymously and irresponsibly; why not put this otherwise noxious trait to good use?

As a symbol of male irresponsibility—and a socially sanctioned symbol at that—one could hardly ask for anything better than artificial insemination with the sperm of anonymous donors. It raises male irresponsibility to the high level of a praised social institution, and it succeeds in getting males off the hook of fatherhood and parenthood in a strikingly effective and decisive way. The anonymity is an especially nice touch; no one will know who did what, and thus there can never be any moral accountability. That is the kind of world all of us have wished we could live in from time to time, especially in its sexual subdivision. From the perspective of the sperm donor, if the child's life turns out poorly, the donor will neither know about that nor inconveniently be called upon to provide help, fatherly help. Home free!

IV. Feminism and Fatherhood

As a movement, feminism has long had a dilemma on its hands. If women are to be free of the undue coercion and domination of males, they must establish their own independent sphere of activities and the necessary social and legal rights to protect that sphere. Women cannot and should not leave their fate in the hands of males, much less their reproductive fates. Meanwhile,

feminists have also deplored feckless, irresponsible males who leave women in the lurch. Yet if males are to be encouraged to act more responsibly, to take seriously their duties to women and children, then they must be allowed to share the right to make decisions in those domains that bear on their activities and responsibilities. Males, moreover, have rights corresponding to their duties; they should be empowered to do that which their moral duties require of them.

For the most part, this dilemma has been resolved by the feminist movement in favor of stressing the independence of females from male control. This is evident in two important respects. First, in the abortion debate there has been a firm rejection of the claim that males should be either informed that a woman is considering an abortion or that the male should have a right to override her decision. The male should, in short, have neither a right to information nor choice about what happens to the conception.

Second, in its acceptance of single-parent procreation and motherhood, for both heterosexual and lesbian women, some branches of feminism have in effect declared fathers biologically irrelevant and socially unnecessary. Since this kind of motherhood requires, as a necessary condition, some male sperm (provided *in vitro* or *in vivo*), it has not been possible to dispense altogether with males. No such luck. But it has been possible to hold those males who assist such reproduction free of all responsibility for their action in providing the sperm. The only difference between the male who impregnates a woman in the course of sexual liaison and then disappears, and the man who is asked to disappear voluntarily after providing sperm, is that the latter kind of irresponsibility is, so to speak, licensed and legitimated. Indeed, it is treated as a kindly, beneficent action. The effect on the child is of course absolutely identical—an unknown, absent father.

Both of these moves seem understandable in the short run, but profoundly unhelpful to women in the long run. It is understandable why women would not want their abortion decision to depend upon male permission. They are the ones who will have to carry the child to term and nurture, as mothers, the child thereafter. It is no less understandable why some women want children without fathers. Some cannot find a male to marry but do not want to give up motherhood altogether; they view this as a

course of necessity, a kind of lesser evil. Other women, for reasons of profound skepticism about males, or hostility toward them, simply want children apart from males altogether.

Please note that I said these motives are "understandable." I did not say they are justifiable. What is short-sighted about either of these choices is that, by their nullification of the moral obligations that ought to go with biological fatherhood, they contribute to the further infantilization of males, a phenomenon already well advanced in our society, and itself a long-standing source of harm for women.

If the obligations of males to take responsibility for the children they have procreated is sharply limited due to women deciding whether to grant males any rights, then males quickly get the message. That message is that the ordinary moral obligations that go with procreation are contingent and dispensable, not nearly as weighty as those of women. For even the most advanced feminists do not lightly allow women who have knowingly chosen to become mothers to jettison that obligation. Mothers are understood to be mothers forever, unlike fathers, who are understood to be fathers as long as no one has declared them free of responsibility. If you are a sperm donor, of course, that declaration can readily be had.

What social conditions are necessary to have the responsibilities of fatherhood taken seriously? The most obvious, it would seem, is a clear, powerful, and consistent social message to fathers: You are responsible for the lives of the children you procreate; you are always the father regardless of legal dispensations; only the gravest emergencies can relieve you of that obligation; you will be held liable if you fail in your duties; and, you will be given the necessary rights and prerogatives required to properly discharge your duties. Only recently has there been a concerted effort, long overdue, to require fathers to make good on child-support agreements. And only recently, and interestingly, has the importance of biological parenthood been sufficiently recognized to lower some of the barriers erected to keep adopted children from discovering the identity of their biological parents, including fathers.

Those feminists who believe that fathers should have no role in abortion decisions should reconsider that position or at least add some nuance. There are probably good reasons to not legally require that fathers be informed that the mother is considering an

abortion; the possibilities of coercion and continuing stress thereafter are real and serious. But that is no reason to dispense with a *moral* requirement that the fathers be informed and their opinion requested if there are no overpowering reasons not to. The fetus that would be aborted is as much their doing as that of the mother, and the loss to the father can obviously be considerable. Acting as if the only serious consequences are for the woman is still another way of minimizing the importance of fatherhood.

Far too much is made of the fact that the woman actually carries the fetus. That does not make the child more hers than his, and in the lifetime span of procreation, childbearing, and child-rearing, the nine-month period of gestation is a minute portion of that span. Only very young parents who have not experienced the troubles of teenage children or an adult child's marital breakup could think of the woman's pregnancy as an especially significant or difficult time compared with other phases of parenthood.

Fathers, in short, have a moral right to know that they are fathers and to have a voice in decisions about the outcome of pregnancy. To deny males such a right is also to reject the very concept of paternal responsibility for one's procreative actions. The right to be a father cannot rest upon someone else's decision to grant such a right; that is no right at all. If the right to be a father is that poorly based, then there will be no better basis for upholding the moral obligations of fathers, or holding them accountable for their actions. I see no possibility of having it both ways. Society often asserts as a general principle that rights entail obligations. In this case, I am arguing the converse: If society wants obligations taken seriously, rights must be recognized.

The argument for a father's moral right to knowledge and choice does not entail a corresponding legal right to force a woman to bear a child against her will. There are a number of prudential and practical reasons not to require legal notification that a woman plans to have an abortion or to require the father's permission. Such a requirement, I suspect, would be both unworkable and probably destructive of many marital relationships. But as a moral norm, this requirement is perfectly appropriate. It puts moral pressure on women to see the need to inform fathers they are fathers, and to withhold such knowledge only when there are serious moral reasons to do so.

Women should, in general, want to do everything possible to encourage fathers to take their role and duties seriously. Women, and the children they bear, only lose if men are allowed to remain infantile and irresponsible. The attempt to encourage more responsible fatherhood and the sharing of childrearing duties while simultaneously promoting the total independence of women in their childbearing decisions only sends a mixed message: Fathers should consider themselves responsible, but not too much; and they should share the choices and burdens of parenthood, but more the latter than the former; and all parents are created equal, but some are more equal than others.

I have mainly laid the emphasis so far on abortion decisions. But the same considerations apply when women, heterosexual or lesbian, make use of donated sperm deliberately to have a single-parent child. Women have been hurt throughout history by males who abandon their parental duties, leaving to women the task of raising the children. A sperm donor is doing exactly the same thing. The fact that he does it with social sanction does not change the outcome; one more male has been allowed to be a father without taking up the duties of fatherhood. Indeed, there is something symbolically destructive about using anonymous sperm donors to help women have children apart from a permanent marital relationship with the father.

For what action could more decisively declare the irrelevance of fatherhood than a specific effort to keep everyone ignorant? A male who would be a party to such an arrangement might well consider himself some kind of altruistic figure, helping women to get what they want. He would in reality be part of that grand old male tradition of fatherhood without tears, that wonderful fatherhood that permits all of the pleasures of procreation but none of its obligations. Women who use males in this way, allowing them to play once again that ancient role in a new guise, cannot fail to do harm both to women and parenthood.

V. Parenthood, Families, and Relationship

A great deal of fun is made these days of those old-fashioned families of the 1950s, especially the television versions, where the emphasis was placed on the family as a unit. They are spoofed in part because they failed to account for all of the families in those

days that were simply not like that. Fair enough. They are derided as well because they often treated the women as empty-headed creatures good for nothing other than cleaning up after the kids and keeping father happy. And sometimes they are attacked because they did not present those fathers as strong leaders and role models for children. Rather, they portrayed fathers as weak and childish, capable of manipulation by wives and children.

But what the old-fashioned families saw clearly enough is that parenthood is a set of relationships, a complex web of rights, privileges, and duties as well as the more subtle interplay of morality in intimate relationships. Feminists have been prone to pose the problem of procreative rights as principally a female problem. Traditionalists have been wont to view fatherhood as a role of patriarchical hegemony. Both are wrong, however, because they fail to see the complexity of the relationship or to place the emphasis in the right place. Both mothers and fathers, as individual moral beings, have important roles as well as the rights and duties that go with those roles.

Those roles, most importantly, are conditioned by, and set in a context of, their mutuality. Each needs and is enriched by the role of the other. The obligations of the one are of benefit to the other; indeed, the mutuality of their obligations amplifies all of them. A mother can better be a mother if she has the active help of a father who takes his duties seriously. Likewise, the father will be a better father with the help of an equally serious mother. The child will, in turn, gain something from both of them, both individually and as a pair. It is important, therefore, that society return fatherhood to center stage not only for the sake of fathers, who will be forced to grow up, but also for mothers, who will benefit from a more mature notion of what fatherhood and parenthood are.

Frail Parents, Robust Duties

Hilde Lindemann Nelson[*]
James Lindemann Nelson[**]

I. INTRODUCTION

The American College of Emergency Physicians recently reported that as many as 100,000 to 200,000 geriatric patients are abandoned each year in emergency rooms across the country.[1] At San Francisco General Hospital alone, three to four such patients are abandoned each week, often suffering from dementia or other ailments that impose serious burdens on their families.[2]

As the human life span is extended in postindustrial societies, those who used to die now grow old and often infirm, yet the younger population, the prop and mainstay of old age, is shrinking proportionately. Care for the aged has begun to be a matter of serious concern, both as a question of public policy and as a familial dilemma. It is also a matter of special concern for women. Since the turn of the century there has been an eight-fold increase in the number of people over the age of sixty-five.[3] In that group, there are sixty-eight men for every hundred women—a ratio that widens to forty-five men for every hundred women by the age of eighty-five.[4] As women tend to outlive men by an average of seven and one-half years,[5] and as "it is now widely recognized that in the United States, families assume a large share of the care-giving responsibility for frail elderly family members,"[6] the problem can be put succinctly: Who should care for Granny?

The answer has typically been other women: Granny's daughter—or daughter-in-law. These are the women Elaine M.

---

* Associate Editor, *Hastings Center Report*.

** Associate for Ethical Studies, The Hastings Center, Briarcliff Manor, New York.

1. Deborah S. Pinkney, *Elderly Straining Emergency Departments*, AM. MED. NEWS, Oct. 12, 1990, at 3.

2. *Id.* at 43.

3. Nancy S. Jecker, *Age-Based Rationing and Women*, 266 JAMA 3012, 3012 (1991).

4. *Id.*

5. *Id.*

6. Suzanne Selig et al., *Ethical Dimensions of Intergenerational Reciprocity: Implications for Practice*, 31 GERONTOLOGIST 624, 624 (1991).

Brody calls "women in the middle," who find themselves rearing children and often working for income while they provide care for their own or their husband's aging relatives.[7] Brody's data indicate that while women want the men in the family to assume equal responsibility for care, men do not do so, although they are willing to provide financial help.[8] The belief that care giving is women's work is so deeply engrained in our culture that it often does not even occur to women that their husbands or brothers could have taken primary responsibility for their parents' care. Brody reports that even women who do not accept traditional views of women's roles often behave as if they do when it comes to frail elderly relatives.[9] A daughter says: "My two brothers and I are all busy lawyers. But when my mother got sick—she lives in another state—my brothers just assumed that I would be the one to fly out to her. And you know something? I did it."[10] Similarly, when a parent-in-law has no daughter, the obligation for care falls on the daughter-in-law rather than the son. "My husband is an only child," said one, "so there was no one else."[11] Said another, when asked how *she* came to be her mother-in-law's care giver, "My husband has two brothers, but he was always the one in the family who took the most responsibility."[12]

The toll this caring takes is sometimes considerable. The physical labor of changing the dirty diapers and wet sheets of incontinent parents, the interrupted sleep so that medications can be given two or three times a night or fears can be calmed, the continual worry that a forgetful parent will leave the stove on or fall and break a bone—these are not the hardest burden to bear.

> The most severe impact of caring for a dependent adult appears to be that it is totally monopolizing and without respite, twenty-four hours a day, seven days a week, 365 days a year. . . . There is gradually isolation of the whole family, but particularly the main caregiver. They no longer go out, no longer invite people over, no longer accept invitations, because

---

7. ELAINE M. BRODY, WOMEN IN THE MIDDLE: THEIR PARENT-CARE YEARS *passim* (1990).
8. *Id.* at 29.
9. *Id.* at 82.
10. *Id.*
11. *Id.*
12. *Id.*

> they cannot leave the dependent person alone and are too nervous about their unpredictable behavior to receive people or to have confidence in substitute care.[13]

Under these circumstances, the care giver experiences extreme mental and physical fatigue. She has difficulty concentrating, cannot sleep, loses marital privacy, and often gives up salaried work or switches to part-time work, despite the repercussions this has on future earnings and pensions.

She does this work of caring because it must be done and no one else in the family is willing to do it. But does she have a moral obligation to do it, or is her work largely supererogatory? The question of what we owe our elderly parents in general has been posed in gender-neutral language, setting the terms for the discussion in ways that importantly falsify it. In our view, the question of filial obligation must not only be recast in terms that acknowledge women's pivotal place in it, but it must also be set within a bioethical framework.

II. WHY FILIAL OBLIGATION IS A BIOETHICAL ISSUE

The question is bioethical for at least three reasons. First, the demographic shift to an aging population with a dwindling number of young people to support it is largely a product of contemporary medicine. The birth-control pill, introduced in the United States in 1960, had a profound impact not only on the birth rate, but on social expectations that family size could be controlled. By the same token, medical interventions increased the human life span. Antibiotics and improved public health measures lessened the threat of infectious disease; the hygienic reform of American hospitals and better prenatal care greatly reduced the danger of dying in childbirth; improvements in anaesthesia and surgical techniques and the large-scale development of corporate pharmacology have greatly contributed to our collective longevity.

In addition to medicine's heavy contribution to the double phenomena of an extended life span and a dwindling youthful population, medicine offers its own answers to the condition of old

---

13. Nancy Guberman, *The Family, Women and Caring: Who Cares for the Carers?*, 17 RESOURCES FOR FEMINIST RES. 37, 39 (1988).

age. As we look to medicine not only to ameliorate the effects of old age, but also to conquer it,[14] we have come to see old age as a disease requiring a medical solution. Viewed from this perspective, the problem of aging in America becomes a problem of health-care rationing and of the legitimate ends of medicine; the question is not only, "Who shall care for Granny?" nor, "How has medicine contributed to the problem?" but, "What kind of care shall Granny receive?"

There is a third way in which the issue of filial duties to the frail elderly is a medical question, and that has to do with the larger issue of the interaction between physicians and patients' families. For a number of very good reasons, medicine's focus is squarely on the patient, and from that point of view families are seen instrumentally, as means to the patient's well-being. Families have a certain epistemic usefulness in that they are assumed to know the former values and preferences of now-incompetent family members, and they have other resources—such as time, energy, money, and emotional support—that are assumed to be at the patient's disposal. When a family member is ill enough to require institutional medical care, the family tends to appropriate this patient-centered, professional ethics for itself. Indeed, medical ethicists teach families that the decisions about medical treatment they are asked to make on behalf of incompetent patients *must* take only the patient's interests into account.[15] Yet when a family is put under ethical pressure to discount the needs of other family members and to ignore the values that give that particular family its own character and identity, it can do serious damage. By honoring the doctor's code of conscience rather than the very different ethics appropriate to families, the intricate web of relationships that make up the household is trampled and sometimes torn. We'll return to these points later, but let us first examine the question of what, if anything, we owe to Granny.

---

14. *See* DANIEL CALLAHAN, SETTING LIMITS: MEDICAL GOALS IN AN AGING SOCIETY *passim* (1987).

15. *See* Linda Emanuel, *PSDA in the Clinic, in Practicing the PSDA*, HASTINGS CENTER REP., Sept.-Oct. 1991, at S6-S7; DaCosta Mason, *On Behalf of the Patient, in Practicing the PSDA*, *supra*, at S9-S10.

III. THE HISTORY OF THE FILIAL OBLIGATION QUESTION

Reasoned accounts of why we do—or do not—owe our parents anything in particular, and what the contents of the obligation—if present—might be, broadly speaking can be divided into four groups. The first group falls under the general category of duties based on indebtedness. The second group is based on consent. The third group of duty theories is based on gratitude. We advocate, however, a fourth account, which consists of a duty theory based on the intimacy of the parent-child relationship. Each of these approaches, and the difficulties they present, are discussed below.

A. Duties Based in Indebtedness

The first group consists of duty theories based on indebtedness: our parents gave us life and took care of us when we needed care; in return, we owe them care when they are in need. Aristotle and Aquinas both held this view,[16] but we will quote Blackstone's formulation:

> The *duties* of children to their parents arise from a principle of natural justice and retribution. For to those who gave us existence, we naturally owe subjection and obedience during our minority, and honor and reverence ever after: they, who protected the weakness of our infancy are entitled to our protection in the infirmity of their age; they who by sustenance and education have enabled their offspring to prosper, ought in return to be supported by that offspring, in case they stand in need of assistance.[17]

---

16. Aristotle stated:

> That is why it would seem that a son does not have the right to disown his father, whereas a father has the right to disown his son. A debtor must pay his debt, but nothing a son may have done (to repay his father) is a worthy return for everything the father has provided for him, and therefore he will always be in his debt.

ARISTOTLE, NICOMACHEAN ETHICS § 1163b, at 244 (Martin Ostwald ed. and trans., 1962). Aquinas believed that since our parents are, next to God, the "closest sources of our existence and development," we owe them respect, reverence, and services. 13 THOMAS AQUINAS, SUMMA THEOLOGICA, ques. 101, art. 1 (Blackfriars ed., 1968).

17. 1 WILLIAM BLACKSTONE, COMMENTARIES *453-54.

Christina Hoff Sommers composes a variation on this theme. She argues that the context of family life in our culture creates certain expectations that children will respect their parents; these expectations have the moral force, if not the specificity, of explicit promises. Sommers writes:

> In promising, the content of the obligation is verbally explicit. But this feature is not essential to the formation of specific duties. In the filial situation, the basic relationship is that of nurtured to nurturer, a type of relationship which is very concrete, intimate, and long-lasting and which is considered to be more morally determining than any other in shaping a variety of rights and obligations.[18]

B. Duties Based on Consent

If the indebtedness theories sound rather quid pro quo, the second group of theories denies the debt. These might be called the I-didn't-ask-to-be-born group. In a well-known essay, Jane English, distinguishing between favors which must be reciprocated because they have been solicited, and voluntary sacrifices, which have not been requested and require no return, categorizes parental sacrifices as voluntary and concludes that if the relationship between adult children and their parents is one of friendship, it is subject to the ordinary give and take of any other friendly relation—if not, not.[19] Mark Wicclair distinguishes between duties of indebtedness and duties of gratitude, but he claims children owe their parents a debt of gratitude only insofar as they value the benefits they received from their parents. There is no duty of indebtedness, he says, because the child did not contract for the goods and services the parents provided.[20] Nancy Jecker too finds an obligation of gratitude for nurture and education under certain carefully specified conditions, but like Wicclair, does not derive a duty of care from this obligation.[21] For this group of

---

18. Christina H. Sommers, *Filial Morality*, 83 J. PHIL. 439, 446-47 (1986).

19. Jane English, *What Do Grown Children Owe Their Parents?*, *in* HAVING CHILDREN: PHILOSOPHICAL AND LEGAL REFLECTIONS ON PARENTHOOD 351-56 (Onora O'Neill & William Ruddick eds., 1979).

20. Mark R. Wicclair, *Caring for Frail Elderly Parents: Past Parental Sacrifices and the Obligations of Adult Children*, 16 SOC. THEORY AND PRAC. 163 *passim* (1990).

21. Nancy S. Jecker, *Are Filial Duties Unfounded?*, 26 AM. PHIL. Q. 73, 78 (1989).

theories, there cannot be a duty to care for frail parents unless the child consents to such care, for all duties are based on consent.

C. Duties Based in Gratitude

The third group of duty theories is less problematic than the contract theories; it is based on gratitude. Joel Feinberg, for one, finds a duty of gratitude that is quite different from the sort of indebtedness Blackstone describes.

> But gratitude, I submit, feels nothing at all like indebtedness. . . . My benefactor once freely offered me his services when I needed them. There was, on that occasion, nothing for me to do in return but express my deepest gratitude to him. . . . But now circumstances have arisen in which he needs help, and I am in a position to help him. Surely, I *owe* him my services now, and he would be entitled to resent my failure to come through.[22]

Jeffrey Blustein argues that gratitude is owed even where the benefactor was obliged to offer the needed services.

> [I]t is the motive of the giver that gratitude regards, not the obligatoriness of the actions. Indeed, since the degree of obligation to gratitude is to be judged in part by how beneficial the service was to the obligated subject, and since the duties of parents are duties to protect and promote the child's serious interests, grown children may actually have more of a duty to show gratitude for benefits that were owed them than for those that were not.[23]

For both authors the duty cashes out practically, in deeds and not just sentiments.

---

22. Joel Feinberg, *Duties, Rights, and Claims*, 3 AM. PHIL. Q. 137, 139 (1966).

23. JEFFREY BLUSTEIN, PARENTS & CHILDREN: THE ETHICS OF THE FAMILY 183 (1982).

D. Difficulties with These Views

The trouble with the duty theories based on indebtedness is that they don't seem to recognize that parents *owe* the children a decent minimum of the goods and services they provide. When parents fail to provide these goods we put the children in foster care and punish the parents; it is illegal and immoral to neglect one's children. The duty model sees the relationship between parents and children as symmetrical—or, for Aristotle and Blackstone, weighted heavily in favor of the parents—but in fact, if the child is owed what the parents provide, why should the child respond with anything more than a thank you? Sommers's version is equally open to this criticism; to avoid begging the question, she must provide us with good reason to regard expectations of filial respect as legitimate, not just as ubiquitous.

But the I-didn't-ask-to-be-born theories are not very satisfactory either. If we reduce the relationship between parent and child to a simple matter of choice—if we see it as purely an *elective* affinity (to use Goethe's term)—we seem to say too little. For one thing, despite birth control and abortion, many births are not a matter of choice, and even when parents choose to have *a* child, they no more choose *this* child than it chooses them. Working out an ethics of the parent-child relationship on the basis of election and consent is a very odd thing to do, as these people did not elect each other and have not consented to the relationship. For another thing, family ties differ importantly from the ties of, say, friendship or collegiality in that they persist over time irrespective of the vicissitudes of life. You are free to allow your friendships to fade by mutual consent, but your sister is always your sister. A model of obligation based solely on affection is too contingent; we can't count on people if they have duties to us only insofar as they like us.

There is a related problem with the claim that we have duties only insofar as we value the *goods* we receive from our parents. If a mother and father have provided excellently for their child, proving themselves to be affectionate, sensible, estimable parents, the child owes them gratitude *whether or not* she values the benefits. Suppose on her birthday a master cabinetmaker presented his daughter with an inlaid and delicately designed writing desk that he had made himself, only to have her inform him that

it was not to her taste, and pass on to her next gift without thanking him. We would say of her not only that she has failed in gratitude, but that she did so because she failed to see the value of the desk and of her father's love for her. She lacks moral perception. In any case, a duty of gratitude will not help us answer the question whether we must care for Granny, because gratitude, unlike indebtedness, cannot be discharged by giving something in exchange for what was received. The daughter cannot present her father with a set of needlepointed chair seats of her own design and then say to herself, "There. Now I needn't be grateful for that desk anymore." Gratitude, we submit, is more like vision than like a bargain: we see, we appreciate, we are moved to thanks. This moment of seeing can occur again and again, when the daughter sits at the desk to write a letter, for example, many years after the gift was given.

E. Our View: A Nonlibertarian Model of the Family

We are the fourth group. We too think adult children have duties of performance to their parents, even though they do not owe them repayment for the goods and services the parents provided when the children were young. But while we are sympathetic to Feinberg and Blustein, we base the duty in another aspect of the parental obligation, rather than in filial gratitude.

With the exception of a very few, short-lived utopian experiments, the standard social arrangement in any culture we know of for protecting, nurturing, and socializing children has been to entrust that care to the family.[24] The division of this caretaking labor between parents has typically been inequitable—indeed, Sara Ruddick calls it maternal work[25]—but the obligation is recognized in common morality and in law. Parents owe their children both protection from harm and those things required for human growth and development, including a reasonable amount of education.

It is also a part of the common morality—and not only in our own culture—that parents owe the child love. That is, parents are

---

24. WILLIAM J. GOODE, THE FAMILY 6 (2d ed. 1982).

25. *See* SARA RUDDICK, MATERNAL THINKING: TOWARD A POLITICS OF PEACE 12, 29 (1989).

to single out their children for special cherishing and affection, not only because loving and being loved is a part of the child's socialization, and not only because our hormones and our own socialization prompt us to find our children lovable, but because the familial work of forging the child's identity is fueled by love; without it the child cannot achieve either the sense of belonging to the family or the sense of separateness from the family that are essential to the child's self.[26] Without an abiding place in the world where the child is cherished and singled out specially from all others, the child is incomplete—like Colin Turnbull's mountain people,[27] she or he grows up a spiritual and social cripple.

1. A Narrative View of Personal Identity

The transaction that takes place between parent and child, where the parents give and the child receives, is, for the child, morally indeterminate when the child is young. By "morally indeterminate," we mean that the transaction has no particular moral significance to the child. There is a parallel here in literature. Consider the opening passage of James Joyce's *Portrait of the Artist as a Young Man*:

> Once upon a time and a very good time it was there was a moocow coming down along the road and this moocow that was coming down along the road met a nicens little boy named baby tuckoo. . . .
>
> His father told him that story: his father looked at him through a glass: he had a hairy face.[28]

We attach no particular meaning to the babble of Baby Tuckoo and the hairy-faced father. The meaning is indeterminate until we are further along in the novel. Then we see that Baby Tuckoo is Stephen himself, the cuckoo set in the wrong nest, and that the hairy-faced father is an image of the Roman Catholic Church, which Stephen, in a recapitulation of Lucifer's rebellion, refuses to serve. So it is with our own childhood: the meaning of

---

26. SALVADOR MINUCHIN, FAMILIES AND FAMILY THERAPY 46-48 (1974).

27. COLIN M. TURNBULL, THE MOUNTAIN PEOPLE *passim* (1972).

28. JAMES JOYCE, A PORTRAIT OF THE ARTIST AS A YOUNG MAN 7 (Penguin Books 1976) (1916).

the events and relationships experienced there is not determined until later in the life story.

When we become adults, it is one of the tasks of moral agency to sift and sort our past as we try to make sense of our experiences. Our family history is perhaps the greater part of the grist for this activity, which we see as largely *narrative*. As we reflect on our past we identify themes, select incidents, and draw connections in order to create meaning for our lives. We bring into the present many of the values and virtues we acquired in childhood, but now we are free to confirm certain of them and reject others. We can do so in the choices we make about our lives—choices about education, friendships, jobs, and so on. In early adulthood, we begin to plot out our life story, using the first chapter our parents wrote with us as the point of departure. That first chapter can now take on meaning as a guide to action, if we integrate the present moment of decision into the overall narrative of our lives, either by ratifying our existing course or by deliberately charting a new one. We need not do so, but it is open to us to make our present choices in such a way as to weave our past into our future, consciously and thoughtfully constructing a meaningful moral autobiography. This process, akin in many respects to what Margaret Urban Walker has called strong moral self-definition,[29] is capable not only of shaping our future, but also, we submit, of transfiguring the past.

It can do so in this sense. The parental giving and filial receiving characteristic of early childhood is a major theme of the very beginning of the child's story, and one cannot yet tell what moral significance the child will make of it. But when that child grows into full moral agency, he is able retrospectively to make that giving and receiving mean a variety of things, depending on how he treats his parents now. If his parents now come to him in need and he spurns them, he is declaring that the relationship he had with them as a child was largely instrumental: he was using them only as a means to his own ends, and they are no more to him than that. Alternatively, if he now responds to their needs, he is *redeeming* that childhood relationship of its instrumentality, and declares by his actions that he was not merely using his parents to provide goods and services for him.

---

29. Margaret U. Walker, *Moral Particularity*, 18 METAPHILOSOPHY 171, 173 (1987).

2. The Inherently Noninstrumental Nature of Intimacy

But why shouldn't he treat that relationship instrumentally? How is it wrong, if his parents were obliged to provide him care when he was young? If Jones owes Smith money and pays the debt, Smith owes him nothing in return. The disanalogy between Smith and the adult child, however, lies in the nature of the goods provided in childhood. For oddly enough, one of those goods is encumbered with a special duty to the giver. Parents, it will be remembered, are obliged to love their children. To discharge this duty, they must put themselves into a relationship of affectionate intimacy with the child—to share, to an appropriate extent, the parental self with the child. The good that is here given is not food or clothing or any other object, but a bond that inherently tends toward mutual affirmation. The parent-child relationship is, in its initial stages of course, not mutual in fact, but it is informed by love, and love seeks mutuality.

In other words, those within a loving relationship cannot remain indifferent to the needs of others within it, for if they do, the relationship ceases to exist. Initially this loving relationship is heavily one-sided, because infants' capacity to respond is limited. But as the child grows within this bond of love, he grows in awareness of the other selves that are also within the bond, for only in doing this can he get the good of love. It is a little like a conversation: I must speak as well as listen, for if I do not I have put myself outside the conversational circle. We might say that the special duty to respond to the needs of those we love is a side effect we have to accept if we are to get the goods of intimacy so vital to our well-being. In giving their children this relationship, parents *encumber* their children, teaching them to be, not unattached, atomic individuals, but people joined to other people in many different kinds of relationships. As they live out what they are teaching, parents become vulnerable to the harm that will occur if the relationship is severed.

Consider how the harm is played out in *King Lear*. The King's mistake is not solely in confounding Cordelia's lack of ceremony with ingratitude, nor is he merely a victim of actual ingratitude at the hands of Goneril and Regan. What is "sharper than a serpent's tooth" is the thoroughgoing manner in which Goneril and Regan destroy the bonds of familial love. They

continue to use the old man as means to their ends long past the dictates of necessity, refusing to acknowledge that the old man, in loving them, has legitimate reason to trust them in turn. The promise of reciprocity inherent in intimacy, fidelity, and the other goods of family life is broken, the family is sundered, the bond between parent and child betrayed, and Lear himself becomes, as Sommers notes, "a shadow." But contrary to Sommers's account, such infidelity is not wrong because there is a widespread expectation that children will keep faith with their parents; it is wrong because the relationship must be encumbered with such fidelity if it is to do its job for the child in the first place.

3. Objections to This View

Those who believe that the basis for all obligation is consensual will at this point argue that it is not possible to bind adults by what was done for them as children, as they were not then moral agents, and in no position to consent to obligations thrust upon them at that time. But this libertarian position fails to acknowledge the extent to which human beings are forced to depend on each other—the extent to which liberal theories of self-determination themselves depend on what has been aptly named the shadow-work of those who are willing to care for dependent and helpless members of society.[30] We do not think a consensual model of obligation is adequate for any community; and certainly it is inadequate for families, where many of the people living in them are needy and dependent. Further, the loving intimacy parents owe their children requires parents to create the conditions for a noninstrumental, mutual relationship, which they must then proceed to fashion. If the adult child repudiates the relationship thus fashioned without good cause, he has betrayed the parents' legitimate expectation that their love will be returned. In doing so he inflicts gratuitous pain on them.

A second objection to our account of filial obligation might be to argue that if an adult child is without concern for his parents, it simply shows that the parents were not successful in creating

30. For a very helpful discussion of this point, see Annette C. Baier, *The Need for More than Justice, in Science, Morality & Feminist Theory* (Marsha Hansen & Kai Neilsen eds.), 13 CANADIAN J. PHIL. 41, *passim* (Supp. 1987).

a loving relationship. If the relationship isn't there, the child can't be said to have betrayed it, and no duties follow. But this is not necessarily the case. A child might come to repudiate the relationship with his parents not because the parents did not set up a loving relationship with the child, but because of selfishness, because the relationship did not meet the child's standards of what love ought to be, because the relationship became burdensome, or because the relationship created a moral wrong greater than the wrong of repudiating the parent (for example, if the frail parent's alcoholism was endangering the adult child's relationship with other members of the family). Each of these possibilities requires its own analysis.

IV. IMPLICATIONS OF OUR VIEW OF THE FILIAL OBLIGATION QUESTION

We earlier characterized care for aged parents as a bioethical problem—an instance of the skewed way in which medicine sees the family. When physicians, prompted perhaps by prevailing currents in ethics, take as an unexamined norm the ideal of the human being as an autonomous moral agent, unencumbered by relationships, cultural ties, or duties to others, they are going to make treatment decisions that reflect this bias. They will create a fiction of a demented person's autonomy, or encourage a patient to decide on a course of treatment without taking into account the familial context within which the treatment is to be provided. Our approach can help here. When the individual's moral agency is seen as nested within a web of intimacy, decision making balances all relevant needs, and not just those of the patient. When there is a conflict, the individual's and the family's process of strong moral self-definition can guide its resolution, as justice and the particular exigencies of the situation dictate.

The parent-child relationship existed long before the child was capable of informed choice. Yet the past exerts a moral pull on us, as it encumbers us with at least a prima facie duty to cherish our elderly parents. The duty is particularistic in that it must be fitted to the parents' needs and the child's means. Moreover, there could be a number of ways to meet a need, all of which are acceptable, but one of which is preferable in a particular case because it best fits the life patterns of those involved.

Whether the need is for nursing care, help with shopping, or financial assistance, the child must respond, but he need not do so single-handedly. Just as public education helps parents discharge their duty to socialize and educate their children, so the larger community can, and does, help with the care of its more vulnerable citizens.

As the social problem of care for the elderly grows more severe, it becomes increasingly important that we achieve a social and political consensus that the burden be distributed more equitably. Filial responsibility laws, for example—which are statutory in one form or another in thirty states—must be rejected on the grounds that they are unfair. These laws are an instance of a pattern characteristic of a liberal state; as Laura Purdy has observed, "the public realm works only if many real human needs are taken care of somewhere else—a place where the individualistic conception of human relations predicated of that realm do not hold."[31] That is, filial responsibility laws relieve the public of its responsibility to the elderly by relegating that responsibility to a private sphere—the family—whose morality, unlike that of the public sphere, is not based on individual interest satisfaction, but rather on self-sacrifice and the interest satisfaction of others. Further, such laws perpetuate a sexually oppressive feature of traditional family life by imposing the burden of care without the possibility of equitable enforcement. Constitutionally protected rights to privacy close the family off from public scrutiny and ensure that inequities in the distribution of caring labor will go unsanctioned.

While filial responsibility laws could help to prevent the worst abuses of the elderly, as for instance abandoning them in emergency rooms, California, where such dumping is on the rise, has a filial responsibility law that seems ineffective in the face of such abandonment. One trouble with the California law is that it specifies no jurisdiction for enforcement of the law; it is not clear which state agency is empowered to bring suit in cases of nonsupport.[32] But even if this flaw in the statute were corrected, we

---

31. LAURA M. PURDY, IN THEIR BEST INTEREST? THE CASE AGAINST EQUAL RIGHTS FOR CHILDREN 64 (1992).

32. Kris Bulcroft et al., *Filial Responsibility Laws: Issues and State Statutes*, 11 RES. ON AGING 374, 385-86 (1989).

would still find it unjust because, where families are commanded to care, women will do the caring. Government agencies cannot avoid such an outcome.

Instead, we propose an analogy to education. We have agreed as a society that it is of fundamental importance to have an educated citizenry, and so the cost of that education is shared by all, even the childless. The needs of the elderly ought to be provided for on the same basis. Where possible, home care for those who could benefit from it would be a cost-effective solution; where home care is not feasible, nursing home or hospital care must be provided. If this care is embedded in a larger reform of U.S. health care that moves to the less wasteful system of a single-party payer and that strives to educate people to accept that we must not provide premier health care to some while denying any kind of care at all to the thirty-seven million uninsured in this country, it would be a distinct improvement over the present state of affairs.

It might be argued that universal, public education is a recent development, and one that became politically feasible only when the electorate saw that it was essential to the maintenance of a liberal nation-state. This is surely a buttressing consideration, but it is not the principle one. In a just society, education is universal and public because it is essential to human flourishing, not because it is necessary to the state. Similarly, care for the elderly is so essential to human flourishing that we cannot in decency fail to provide it, even if there is no direct benefit to the state.

But to say all this is not to lose sight of adult children's responsibility; it is merely to say that they need help in meeting it. Men in particular must do more. They have come to take an increasingly larger share of the care of their children; they can also be educated to take greater responsibility for other kinds of care. The workplace too needs to be restructured to allow us to meet our obligations. This takes time and social pressure, as women have discovered over the last twenty years or so, but in the face of the large amounts of care that will be required as the demographic shift toward an aging society becomes even more pronounced, we cannot afford to give up on this.

State and federal policies regarding care of the old too often assume that the family is a handy and inexpensive solution to the

problem. It is not. But as we evolve equitable and humane mechanisms for meeting the needs of the elderly, the family's place in all this needs to be examined intelligently and imaginatively. We all have a large stake in the issue: we will all (with any luck) be old ourselves one day. It is our own good that we are trying to secure, and our own disaster that we must try to avert. If we do not reshape the social structures that speak to the frailties of our parents and grandparents, we may one day ourselves be abandoned to the indifferent mercy of strangers.

Intergenerational Responsibility and Family Obligation: On Sharing

Lee E. Teitelbaum[*]

I. INTRODUCTION

The question of what family members owe to each other has become acute for a number of reasons: some demographic; some economic; and some social. The issue, of course, is an old one. During the sixteenth century, support for those who were disabled through youth (typically orphans and abandoned children), age, or infirmity was first a matter for private charity and, when that failed, became a matter of local responsibility.[1] However, the cost of local responsibility was soon recognized, and recognition was swiftly followed by efforts to minimize the public obligation by, for example, punishing those who were "voluntarily" disabled —vagrants and beggars[2]—and shifting to families the primary obligation for support of their disabled relatives.[3]

Although these so-called Poor Laws were adopted in the colonies[4] and have remained on the books in many states, there is little evidence of their widespread use. Public involvement with

---

* Dean and Professor of Law, University of Utah College of Law. The author would like to express the deepest gratitude to Ronald Boyce, Thomas Chancellor, Leslie Harris, Bonnie Mitchell, Linda Smith, and Ned Spurgeon for their thoughtful comments on a draft of this Article, and to Ms. Becky Parr and Ms. Jonell Thomas, for their creative research assistance. The author alone is responsible for any error or infelicity.

1. *See generally* Stefan Riesenfeld, *The Formative Era of American Public Assistance Law*, 43 CAL. L. REV. 175 (1955).

2. *See, e.g., An Act for Punishment of Rogues, Vagabonds, and Sturdy Beggars*, 39 Eliz., ch. 4 (1597) (Eng.).

3. The Elizabethan Act of 1601 for the Relief of the Poor, 43 Eliz., ch. 2 (1601) (Eng.), provided, among other things, "[t]hat the father and grandfather, and the mother and grandmother, and the children of every poor, old, blind, lame and impotent person, or other poor person not able to work, being of a sufficient ability, shall, at their own charges, relieve and maintain every such poor person." *Id.* § 7. The effect of this legislation was to place primary responsibility for support upon the close relatives of poor persons, with public responsibility arising only secondarily.

4. *See* WALTER I. TRATTNER, FROM POOR LAW TO WELFARE STATE: A HISTORY OF SOCIAL WELFARE IN AMERICA 15-19 (4th ed. 1989). American statutes followed the pattern of the English poor laws, placing primary responsibility for support on near relatives of the disabled person.

care of the needy was also scant, particularly during the late eighteenth and nineteenth centuries.[5] Where public assistance was available, it largely took the form of institutional care, particularly for the destitute elderly who were commonly placed in public facilities. Assistance for those who remained in the home was rarely available except from private sources.

With the first third of this century came increased interest in public programs, primarily for children[6] but also for disabled workers. Workers' compensation statutes, although primarily intended to protect employers against large liability awards, provided a limited scheme of protection for employees as well. The effects of the Great Depression substantially accelerated the movement toward public programs while concurrently depleting individual and private resources. Economic collapse limited not only the capacity of citizens to care for themselves, but also the capacities of other private individuals and groups to do so. Private charitable agencies disappeared almost as quickly as private entrepreneurs.[7]

The 1930s, accordingly, saw the creation of the modern social security and welfare system and, with it, the first substantial, direct participation of the federal government in care for the disabled and elderly. Concomitantly, reliance on private and familial responsibility declined.[8] To some considerable extent, these new programs were regarded as a kind of insurance, purchased during employment by workers and their employers and drawn against upon retirement or disability. To the extent that social welfare was need-based, the theory was of societal and intergenerational responsibility rather than familial responsibility.

---

5. *See generally id.*

6. The impetus for these reforms came from the 1909 Conference on Dependent Children which resulted, three years later, in the creation of the United States Children's Bureau. Among the many sources on Progressive era "child-saving" is ANTHONY M. PLATT, THE CHILD-SAVERS: THE INVENTION OF DELINQUENCY (1969). Although much of the response to fears of family disintegration associated with child-saving emphasized institutional care, the "mother's pension"—providing cash assistance to mothers with dependent children—also appeared between 1911 and 1935. *See* TRATTNER, *supra* note 4, at 201-02.

7. About one-third of the nation's private agencies collapsed during the period from 1929 to 1932. TRATTNER, *supra* note 4, at 249.

8. Although one can quarrel with any "watershed," the Federal Emergency Relief Act of 1933 provided one significant shift towards public protection for the unemployed and disabled. *Id.* at 259.

The currently employed would support the elderly or disabled through federal and state taxes, expecting similar assistance from the rising generation when their time of need arrived. And as long as federal support for health and social welfare programs remained substantial, family responsibility laws largely remained dormant.[9]

Conditions, however, have changed. Some of the changes are demographic. The aging of the population is old news.[10] The population in need of support—the very elderly and at least partially disabled—has grown remarkably because advances in health care have extended life spans.[11] Other changes are social in nature. Demographics may describe the larger number of aged citizens, but divorce policy (coupled with differences in remarriage rates for women and men) at least partly explains the larger proportion of single elderly persons (particularly single women) for whom support is especially needed.[12] And some of the changes are economic. On the one hand, social security entitlements have increased since the 1970s, reducing significantly the proportion of elderly persons below the poverty level. At the same time, those who are poor are also highly vulnerable. Federal contributions to federal-state programs such as Medicaid have been restrained,

---

9. *See* George F. Indest III, *Legal Aspects of HCFA's Decision to Allow Recovery from Children for Medicaid Benefits Delivered to Their Parents Through State Financial Responsibility Statutes: A Case of Bad Rule Making Through Failure to Comply with the Administrative Procedure Act*, 15 S.U. L. REV. 225, 253-57 (1988).

10. A white male who lived to age 20 in 1900 had a life expectancy of 62.2 years; a white male at age 20 in 1980 has a projected life expectancy of 72.7 years. John H. Langbein, *The Twentieth-Century Revolution in Family Wealth Transmission*, 86 MICH. L. REV. 722, 742 (1988). A 20 year old white female had a life expectancy of 63.8 years in 1900, but of 79.7 years in 1980. *Id.* These are gains of more than 10 years for men and almost 16 years for women. *Id.* The increase for non-whites is lower but still substantial.

11. The elderly are, of course, heterogeneous. Relatively few in the 65-74 year age group suffer disability, but the frequency of disability increases substantially with increasing age. NORMAN DANIELS, AM I MY PARENTS' KEEPER?: AN ESSAY ON JUSTICE BETWEEN THE YOUNG AND THE OLD 8 (1988). One study reveals that personal care assistance is required by only about 7% of those between 65-74, by 15% of those between 75-84, and by 44% of those aged 85 and older. *Id.* at 8 n.8.

12. While increases in men living alone have also occurred, those increases are largely found in younger rather than older age groups. Older women seem less likely to remarry than older men. *See id.* at 9. Nine percent of men aged 55-74 were widowed or divorced in 1970, compared with 34% of the women in that age group. Frances E. Kobrin, *The Fall in Household Size and the Rise of the Primary Individual in the United States*, *in* THE AMERICAN FAMILY IN SOCIAL-HISTORICAL PERSPECTIVE 100, 104 (Michael Gordon ed., 3rd ed. 1983).

while health care costs have risen enormously.[13] Under these circumstances, and impelled by conservative emphases upon limiting federal (and other governmental) roles in "private" areas of concern, there is once again interest, largely at the state level, in substituting family for local responsibility.[14]

This Article addresses several questions related to intergenerational family responsibility, emphasizing obligations from children to parents.[15] The first section examines briefly the importance of family decisions and wealth distributions in modern society. Section two addresses two ethical questions arising because families devote much of their loyalty and resources to their members. One is whether it is right for persons to distribute wealth to family members rather than to others—an issue that has generated substantial controversy. The other ethical question supposes that such preferences are justifiable and concerns the

---

13. Health care expenditures in the United States accounted for more than 11% of the Gross National Product in 1987, compared with a 1.7% share in 1945. B.K. Atrostic & Sonia Conly, *From Baby Boom to Baby Bust: Tax Implications of Changing Demographics*, *in* NATIONAL TAX ASSOCIATION—TAX INSTITUTE OF AMERICA, PROCEEDINGS OF THE EIGHTY-SECOND ANNUAL CONFERENCE 177, 177-78 (Frederick D. Stocker ed., 1989). In addition, over the last two decades, expenditures on health grew more rapidly than the population and the cost of health care rose more rapidly than the overall price level. *Id.*; *see also*, Robert Whitman & Diane Whitney, *Are Children Legally Responsible for the Support of their Parents?*, 123 TR. & EST. 43, 44 (Dec. 1984).

14. The two most often discussed examples are family responsibility laws at the state level and the reliance on such laws in seeking family contributions to federal-state benefit programs, particularly Medicaid. *See generally*, Catherine D. Byrd, *Relative Responsibility Extended: Requirement of Adult Children to Pay for Their Indigent Parents' Medical Needs*, 22 FAM. L. Q. 87 (1988); Indest, *supra* note 9; Renae R. Patrick, *Honor Thy Father and Mother: Paying the Medical Bills of Elderly Parents*, 19 U. RICH. L. REV. 69 (1984); Whitman & Whitney, *supra* note 13.

This movement may gain support from public concern, arising partly from media coverage about "granny dumping": the practice of abandoning elderly relatives in hospitals or other public places. A survey by the American College of Emergency Physicians reported that up to 70,000 elderly persons were abandoned by family members in 1991. *Generation Squeeze: Families Abandoned 70,000 Elderly in 1991, Group Says*, SALT LAKE TRIB., March 27, 1992, at A1.

15. Although not within the scope of this discussion, parental obligations to children are also due for some reconsideration. Here, too, social, legal, and economic conditions have changed over the last twenty or thirty years. Until divorce became a national pastime, the configuration of parent-child interests was relatively simple. In most cases, children had formal social obligations to only one set of parents and parents had social and legal obligations to only one set of children. More complicated parent-child relations are among the obvious implications of a divorce rate that will, for the generation that married in the 1970s, reach almost 50%. ANDREWS J. CHERLIN, MARRIAGE, DIVORCE, REMARRIAGE 24-25 (1981).

fairness of not only permitting but requiring persons to contribute specially to the well-being of parents in need.

Section three explores the imposition of family obligation from a different perspective. Even if imposition of filial obligation is morally justified, its translation into legal rules presents special difficulties in the context of family life. These difficulties, long reflected in the reluctance of courts and legislatures to regulate family law directly, seem equally troublesome in the context of filial responsibility laws.

The final section suggests a way to approach the care of disabled family members that does not rely on legal rules to shift family responsibility from a public to a private sector. It assumes, rather, that this responsibility is shared by families and other social institutions that control the delivery of health services. The discussion also suggests that analysis of the policies and practices of those institutions provides a better vehicle for defining the nature and extent of family responsibility than do rules addressed to the family alone.

II. THE IMPORTANCE OF FAMILY DECISIONS

The importance of family responsibility is worth a few words because families are commonly thought to play only a minor role in modern life. American sociology[16] and modern social history[17] alike emphasize the limited function of the family in industrial and post-modern society. Households are no longer the principal

16. The shrinking role of the family has been a consistent theme of American sociology. The Chicago school in the 1920s emphasized the increasing isolation of the nuclear family as a consequence of urbanization and, particularly, the loss of old kinship and communal supports. The structuralist-functionalist school after World War II, of whom Talcott Parsons is the most prominent theorist, likewise accepted a loss of function, although his emphasis was more positive. TALCOTT PARSONS ET AL., FAMILY, SOCIALIZATION AND INTERACTION PROCESS (1955). A central aspect of modernization is institutional differentiation, as functions originally discharged by one institution are distributed among several. In Parsons' view, the family had lost educational and economic functions, but this loss freed the family for new, and primarily affective, functions. *Id. See generally*, BRIGITTE BERGER & PETER BERGER, THE WAR OVER THE FAMILY 10-13 (1983).

17. The creation of a separate "domestic" sphere for the family, reflecting both modern economic conditions and the need for providing a sense of equality for women, is a consistent theme in modern social history. *See, e.g.*, John Demos, *Images of the American Family, Then and Now*, *in* CHANGING IMAGES OF THE FAMILY 43 (Virginia Tufte & Barbara Myerhoff eds., 1979).

agents of economic production, education of children, or social control, as they seem to have been during the early history of this country. Those functions have largely been assumed by other social systems: by a legal system that emphasizes individual citizenship and has sought to dissociate rights from other social institutions; by industrialized means of production and national markets; by an economy in which employment by others has replaced land or family business as a source of wealth; and by publicly regulated educational institutions. The family, it is said, has little importance or function, other than in the affective realm, as a source of emotional support for its members.[18]

This sketch seems, however, unduly minimalist. The family plays a central role in modern society, in a quite literal sense. Intermediate administrative and regulatory systems, such as the educational and medical care systems, do not entirely occupy the fields once primarily committed to family authority. Indeed, to a considerable extent, their activities remain dependent on family decisions. More generally, families still serve in significant ways as economic systems, moral systems, and educational systems. They remain, in short, important agencies for the distribution of social goods.

Take, for example, the family's economic functions. Although wealth is no longer produced within the home, and public and private corporate employers largely control the means of production, family members retain considerable control over the generation and consumption of wealth. True, the nature of this wealth has changed. It is no longer found in farms and cottage industries, but in employment and streams of income during and after employment.[19] However, family members still have much to say about the generation of this "new property."[20] Husbands and

---

18. This summary, and much of the following, draws on Lee E. Teitelbaum, *Placing the Family in Context*, 22 U.C. DAVIS L. REV. 801, 813-18 (1989).

19. Professor Glendon observes that "for the majority in modern welfare states, old property (in the sense of traditional assets of real or personal property) is less important than individual earning power and public or private benefits based on such labor. To the extent that there are savings apart from home equity in a middle-aged middle-income family, they tend less to be represented by bank accounts or tangible assets than by employment-related pension plans, profit-sharing plans, insurance or other benefits . . . . Indeed, according to Peter Drucker's estimate, pensions may be the largest single asset of the middle-aged American. . . ." MARY A. GLENDON, THE NEW FAMILY AND THE NEW PROPERTY 93-94 (1981).

20. The phrase new property is most strongly associated with a series of writings by

wives decide how many family members will work, and in what settings. That decision in turn defines the level of consumption available to them and other family members, especially their children.[21] Spouses also typically decide how many children they will have, a decision that has an even greater effect on what is available to those children.[22] When family members generate excess wealth, they may allocate those resources across an almost unlimited range of choices. If children are born or adopted, the family faces distributive choices with respect to education, even though public schooling, from grammar school through college, is available in every state. Across socioeconomic lines, children of wealthy parents are more likely than are poor children to have access to private secondary schools, to the more prestigious private colleges, and to graduate and professional education.

Even within socioeconomic groups, families possess a considerable range of educational choice. Those seeking upward mobility for their children may make what seem to others uncalled for "sacrifices" to provide opportunities for all or some of their offspring. On the other hand, those—like the Amish—who reject customary measures of success may withdraw their children from public education at an early age.[23] It is widely accepted that these educational decisions constitute an important form of intergenerational wealth transmission in modern American society.[24]

Charles Reich. *See* CHARLES A. REICH, THE GREENING OF AMERICA (1970); Charles A. Reich, *Individual Rights and Social Welfare: The Emerging Legal Issues*, 74 YALE L.J. 1245 (1965); Charles A. Reich, *The New Property*, 73 YALE L.J. 733 (1964). While Reich used the phrase to refer to social entitlements, it is now often used more broadly to refer to interests, often contingent, based on employment relations.

21. According to Epenshade's study of parental expenditures, whether the mother works inside or outside the home has a substantial effect on the amount spent on a child. THOMAS J. EPENSHADE, INVESTING IN CHILDREN: NEW ESTIMATES OF PARENTAL EXPENDITURES (1984). Work within the home reduces expenditures by almost 20%. *Id.* at 44-45; *see also* EDWARD P. LAZEAR & ROBERT T. MICHAEL, ALLOCATION OF INCOME WITHIN THE HOUSEHOLD 88 (1988) (expenditures on children tend to increase with the proportion of adults in the household who are employed). It should be noted, however, that the relation between increased expenditures and consumption is not direct. Some part of the expenditures are attributable to the purchase of child care.

22. Epenshade reports that increasing the number of children in a prototypical middle class American family from one to three reduces expenditures per child by 35%. EPENSHADE, *supra* note 21, at 3. Lazear and Michael also find that having an additional child lowers expenditures per child. LAZEAR & MICHAEL, *supra* note 21, at 96.

23. *See* Wisconsin v. Yoder, 406 U.S. 205, 207-13 (1972) (Amish parents withdraw children from public school after eighth grade to preserve traditional Amish way of life).

24. *E.g.*, Langbein, *supra* note 10, at 750-51 ("The modern expectation is that for

Nor does intergenerational distributive choice end with the education of children. Most parents leave their accumulated wealth to their children either directly, when there is no surviving spouse, or indirectly through a surviving spouse who will, it is assumed, provide for the younger generation.[25]

Family wealth also affects the level of health care available to family members. Private and public medical insurance often cover only a part of the cost of serious illness and only some kinds of care. This is true for support of children and for support of adults. A reasonable estimate is that when all types of medical expenses for older Americans are totalled, Medicare pays for only about forty percent.[26] Private insurance coverage, where in effect, is also limited.[27] Many policies include a co-payment of, perhaps, twenty percent and often exclude entirely certain routine health care costs as well as forms of treatment regarded as experimental.

Accordingly, the wealth available to families may determine whether some treatments are provided at all, where care is received, and what kinds of care can be purchased. And nobody is very surprised that scarce organs are sometimes allocated for transplant according to the wealth of the recipient.

Family members in fact supply, either by financial contributions or by personal services, many of the health care needs unavailable from public and private insurance programs. The importance of those contributions can hardly be overstated, but some sense may be conveyed by the fact that families "currently provide about eighty percent of all home health care to the partially-disabled elderly," at considerable cost in terms of money, time, and stress.[28]

---

middle-class wealth, the main intergenerational transfer will occur in mid-life, in the form of educational expenditures. The characteristic wealth of later years, the income streams from the public and private pension systems, do not give rise to heirship.").

25. *See, e.g., id.* at 736.

26. JOSEPH L. MATTHEWS, SOCIAL SECURITY, MEDICARE AND PENSIONS 7:2 (5th ed. 1990). The balance is paid for by private insurance, individual contribution, or (in the case of the very needy) Medicaid. *Id.*

27. Private health insurance covered about 31% of personal health care expenditures in 1987. BARRY R. FURROW ET AL., THE LAW OF HEALTH CARE ORGANIZATION AND FINANCE 197 (1991).

28. DANIELS, *supra* note 11, at 6-7.

III. THE ETHICS OF FAMILY SUPPORT

The fact that family members often devote their resources and loyalty to each other rather than to non-family members, and the importance of those special distributions, presents a complex ethical question. On the one hand, discussions of intergenerational responsibility point to these preferences as evidence of a "natural" order of things which can appropriately be employed as a basis for family responsibility laws.[29] On the other hand, commentators from a variety of perspectives regard family loyalty as a serious social and ethical problem.

A. Ethics and Recognition of Family Loyalty and Support

The desirability of intimate family relations, with the special loyalty and special distributions of resources they may generate, is by no means uncontroversial. Marxists, for example, view the bourgeois family as a vehicle for the accumulation and transmission of capital. In this capacity, it serves not only to reproduce capitalism but to deny authentic liberty to wives and children who are dependent on the income-producing male family head.[30] Habermas, following this view, rejects the conventional portrait of the family as a voluntary, affectionate, inner realm following its own noninstrumental program in favor of an interpretation that emphasizes the conjugal family's function in reproducing capital and its assertion of paternal authority.[31]

For his part, John Rawls, representing ethicists who regard fairness of opportunity as a fundamental aspect of justice, notes that "[t]he consistent application of the principle of fair opportunity requires us to view persons independently from the influences of their social position. . . . It seems that even when fair opportunity . . . is satisfied, the family will lead to unequal chances between individuals. Is the family to be abolished then?"[32] Rawls seems

---

29. *See infra* text accompanying notes 45-50 (discussing tradition of loyalty and support for parents).

30. *See* ELI ZARETSKY, CAPITALISM, THE FAMILY, AND PERSONAL LIFE 90-97 (1976).

31. JURGEN HABERMAS, THE STRUCTURAL TRANSFORMATION OF THE PUBLIC SPHERE 47 (1989).

32. JOHN RAWLS, A THEORY OF JUSTICE 511 (1971) (citations omitted).

to incline in that direction[33] but ultimately concludes, without much discussion, that the theory of justice as a whole does not make doing so urgent.[34]

Michael Walzer likewise notes that because the family views itself as a special group, radically separated from the general society, it is a perennial source of inequality.[35] Accordingly, he suggests, "the simplest way to simple equality, is the abolition of the family."[36]

The importance of educational and health care opportunities such as those mentioned above, which are heavily controlled by the family, demonstrates why support for that institution is difficult to reconcile with the principle of equality. The custom of leaving accumulated wealth to family members aggravates and perpetuates this inequality.

Nonetheless, abolition of the family, or even substantial curtailment of its economic significance, has not been widely advocated. The consequence of the special affective relations within the family—a preference for supporting, assisting, and advantaging family members—has sometimes been accepted as a tolerable qualification on the principle of fair opportunity.[37] Preference for family over others has also been treated as an independent principle of justice.

It may be argued, for example, that family rights are justified as an extension of individual rights of family members to self-determination. Charles Fried suggests "that the right to form one's child's values, one's child's life plan and the right to lavish attention on that child are extensions of the basic right not to be interfered with in doing these things for oneself."[38] "Parenthood," he concludes, "is a kind of physical continuity, a physical continuity which is also bound up with spiritual and moral continuity

---

33. Observing that "the principle of fair opportunity can be only imperfectly carried out, . . . as long as the institution of the family exists." *Id.* at 74.

34. *See id.*

35. MICHAEL WALZER, SPHERES OF JUSTICE 229 (1983).

36. *Id.*

37. Rawls observes that no serious inequalities are likely in the context of an otherwise just society. RAWLS, *supra* note 32, at 511.

38. CHARLES FRIED, RIGHT AND WRONG 152 (1978). In the case of parents, he rests that right upon the "special bond between parent and child based on the facts of human reproduction" which produce a special social bond. *Id.* at 153.

through our influence on our children. The family as an institution expresses these aspects of the parents' personality."[39]

From a utilitarian perspective, a case can also be made that loyalty and altruism are important aspects of human nature which are especially likely to be expressed and effective within the family. Most relations in an industrialized society do not call for such commitments and the anonymity of modern society makes them unlikely. Accordingly, it is important to recognize the domestic sphere as an appropriate and peculiarly important locus for expressions of loyalty and concern for others.

It might further be suggested that the functioning family is, at least within this society, a special relationship. Whereas traditional rights theories emphasize the situation of an individual against the claims of anonymous others, family members do not and cannot suppose any such individualism and anonymity, especially across generations. Indeed, it is not possible to talk about parents without at the same moment having children in mind and vice versa. One term entails the existence of the other, in an ordinary case. Moreover, the identities of parents and children are bound up in such complex ways that it is hard to regard either as an interloper on the other's territory. From this special relationship, perhaps, one might justify a special sense of felt responsibility and obligation within families.

B. Ethics and Imposition of Family Responsibility

For present purposes, then, let us assume that a case can be made for recognizing at least some special distributions of loyalty and resources that prefer other family members. On that assumption, it is next important to consider the appropriateness of legal rules expressing those "natural" preferences as obligations. In this section, our concern is with the question of obligation itself: Is it right to say that family members may not only properly feel and express some special sense of loyalty to their relatives but may be required to make special distributions to assist relatives in need?

---

39. *Id.* at 155. Montaigne comes to mind, observing that "[w]e do not marry for ourselves, whatever they may say; we marry as much, or more, for posterity, for the family. The custom and interest of marriage concern our stock, long after we are dead." MICHAEL DE MONTAIGNE, ESSAYS OF MONTAIGNE, book III, ch. 5, 741-42 (trans. 1990).

1. The Family as an Economic Unit

One obvious set of arguments for charging the cost of financial assistance against familial rather than public resources draws on the familiar notion of the family as an economic unit. As we have seen, the family does function as an economic unit to some considerable degree; indeed, more than is commonly recognized.[40] It may therefore seem appropriate for its members to direct their resources to each other when the need arises.

However, families are least likely to maintain substantial economic relations at just the time support for parents is needed. The parent-child financial relation rarely continues in a substantial way after the children reach majority.[41] As a matter of the production of wealth, family farms and businesses are rare; children who are able to and wish to follow their parents in those enterprises are rarer still.[42] The sources of wealth for family members, therefore, are usually entirely independent of each other.

Nor does the family often continue as a unit of consumption past the adulthood of its younger members. Adult children do not generally continue to live with their parents, even if unmarried. The continuing decrease in family and household size, coupled with the growth of what the Census Bureau unattractively calls "primary individuals"—household heads living alone or with

---

40. *See supra* text accompanying notes 19-28 (discussing family's economic function).

41. Fixing the age of majority is, of course, a complex question. It is not a matter of actual co-residence, since many parents continue to support their children while in college, and it is surely not measured by statutory ages of majority. The first full-time job might provide a better measure.

42. Of course, some instances of nepotism have been culturally normal and even approved, particularly if undertaken with a modicum of discretion. As Joan Wexler observes: "In industrial nations during the nineteenth century the sons of entrepreneurs were expected to succeed their fathers at the helm of the family business." Joan G. Wexler, *Husbands and Wives: The Uneasy Case for Antinepotism Rules*, 62 B.U. L. REV. 75, 76 (1982). The leadership of the Ford Motor Company, for example, passed from Henry to Edsel to three grandsons, the last having the good judgment to refuse to discuss nepotism because "'it would not be in good taste.'" *Id.* at 76 n.10. In governmental settings, filial nepotism has not been as generally accepted, although it is hardly unknown. However, attitudes towards nepotism even in the corporate world have changed somewhat in recent years. *See* Keckeisen v. Independent Sch. Dist., 509 F.2d 1062, 1065-66 (8th Cir. 1975), *cert. denied*, 423 U.S. 833 (1975) (upholding school board policy prohibiting employment of husband and wife at same school in administrator/teacher relationship).

unrelated individuals—reveals the economic and residential separation of family members from each other.[43] Consequently, at the time parents need their children's support, the children will themselves be adults, living in their own households often with their own children, and there will be no substantial economic relationship between them and their parents.[44]

2. *Cultural Bases for a Filial Support Obligation*

If the enforcement of a family support obligation cannot be justified on the basis of the economic and social relations of families at the time such support is usually needed, it might rest on other traditional grounds. Requiring us to support our parents seems, for example, to express a strong cultural tradition, either in a historical or in some non-historical sense.

The historical tradition would justify such laws as reproductions of a happier world only recently lost—a world where several generations live together, exchanging roles of nurturance and support in a natural way. Family responsibility statutes reflect, accordingly, only a return to previously existing norms that have wrongly been abandoned.

Perhaps unfortunately, however, the historical period to which return is sought is more remote than we imagine. Extended families—three or more generations living in the same household—have been rare in Western Europe for many centuries and

---

43. Kobrin, *supra* note 12, at 100.

44. The three generation family does, of course, reappear from time to time, in some guise. Elderly parents do sometimes live with their children, although this seems not to be common and not the preferred arrangement. *See* Andrew V. Winter & Thomas K. Burch, *Values, Perceptions, and Choice in Living Arrangements of the Elderly, in* CRITICAL ISSUES IN AGING POLICY 180, 189 (Edgar F. Borgatta & Rhonda J.V. Montgomery eds., 1987). The same seems to be true of children. *See* Edgar F. Borgatta & Rhonda J.V. Montgomery, *Aging Policy and Societal Values, in* CRITICAL ISSUES IN AGING POLICY 7, 21 (Edgar F. Borgatta & Rhonda J.V. Montgomery eds., 1987) (indicating considerable evidence that children would prefer to purchase, rather than provide directly, needed care for their elderly parents). And divorced children—with their own children—sometimes return, for a period, to live with their parents. However, both social and economic relations between the generations look very different than in the nuclear family. Those relations have already been renegotiated upon the adult child's departure from the household, moving from dependency to something like parity or friendship, and the reconstituted living arrangement has its own character. *See generally* COLLEEN L. JOHNSON, EX FAMILIA: GRANDPARENTS, PARENTS, AND CHILDREN ADJUST TO DIVORCE (1988).

never were common in America. Late marriage, late childbearing, early mortality, and geographically diffuse employment patterns after industrialization all conspired to reduce the incidence of three generation families in general and shared residence in particular.[45]

Moreover, even had that happy past existed, family obligations would look far different now. The number of children available to support their parents has declined and the number and proportion of older persons—and of single elderly women particularly—has increased greatly over the last century.[46] Of course, the relatively few available adult children themselves also have their own family responsibilities and, often enough, the multiple-family responsibilities that go with dissolution of their own nuclear groups.

An alternative source of tradition might be found in norms of family loyalty embedded in our culture. We need not look far for such norms. The Fourth Commandment enjoins children of any age to "Honor Thy Father and Thy Mother."[47] Virgil and Dante celebrate pious Aeneas for carrying Anchises, his father, on his shoulder as he traveled to Rome. Aeneas is described as "pious" because he carried in his father's person a respect for family and tradition.[48] Shakespeare has Lear complain, "[h]ow sharper than a serpent's tooth it is to have a thankless child."[49]

---

45. MICHAEL MITTERAUER & REINHARD SIEDER, THE EUROPEAN FAMILY: PATRIARCHY TO PARTNERSHIP FROM THE MIDDLE AGES TO THE PRESENT 151 (1982); *see also* Carl N. Degler, *The Emergence of the Modern American Family*, *in* THE AMERICAN FAMILY IN SOCIAL-HISTORICAL PERSPECTIVE 61, 63 (Michael Gordon ed., 3rd ed. 1983) (indicating marital ages in Europe were much higher than was true for rest of world and percentage of never married persons was also higher in Western Europe). And during the period of industrialization, especially in the 19th and 20th centuries, employment circumstances increased the extent to which children would establish homes apart from their parents.

Similarly, the extended family was never common in America. *Id.* at 62. Relatively late marriage ages were the rule at least by the time of the American Revolution which, combined with mortality, diminished the opportunities for extended family households.

46. From 1890 to 1930, the increase in the proportion of older single persons due to aging was offset by increases in the survivorship of spouses. Since 1930, however, the proportion of elderly single persons, and especially women, has increased dramatically and particularly in relation to available family to provide support. The ratio of daughters aged 35-44 to "no longer married mothers" has fallen from 2.8 daughters per mother to 1.2 daughters per mother in 1974. Kobrin, *supra* note 12, at 109.

47. *Exodus* 20:12 (King James).

48. I am grateful to Ms. Mary Jane Rosenzweig for this reference.

49. WILLIAM SHAKESPEARE, KING LEAR act I, sc. 4.

However, despite the strength of these injunctions to family loyalty, our modern sense of this tradition is surely a complicated one. As a matter of social theory, we have largely abandoned any strong and direct support for parental control over the decisions of their adult children and, indeed, much of our confidence in parental decisions. Although the nineteenth and twentieth centuries celebrated the "privacy" of the family, they simultaneously saw family failure as a common cause of deviance[50] and family inadequacy as a barrier to the development required of children in the new republic.

The establishment of common schools for poor children during the early part of the nineteenth century, followed by the general adoption of compulsory education throughout the century, substituted public for family responsibility as a way of addressing the juvenile crime problem that was largely attributed to the inadequacy of poor (often immigrant) parents to rear their children appropriately.[51] Progressive educational theory, for its part, emphasized the importance of professional instruction that would allow children not merely to replicate, but to improve and leave behind their parents' world.[52]

Current social and psychological theory likewise insist on the importance of the separation of children from their parents. While the relation between parents and children during infancy is one of personal authority, the value of parental control weakens as the child grows older and a conflicting value arises. Children must not only be kept safe and socialized to accept authority, but they must also develop a capacity for autonomous action. A child who does not learn to make choices within our cultural framework is plainly unable to perform the adult role in society. Indeed, the emphasis on acceptance of authority that is valued during infancy gives way to a normative expectation that children will assert some degree

---

50. *See generally* Lee E. Teitelbaum & Leslie J. Harris, *Some Historical Perspectives on Governmental Regulation of Children and Parents*, *in* BEYOND CONTROL: STATUS OFFENDERS IN THE JUVENILE COURT 1 (Lee E. Teitelbaum & Aidan R. Gough eds., 1977); Teitelbaum, *supra* note 18.

51. Teitelbaum & Harris, *supra* note 50, at 18-20; *see also* U.S. BUREAU OF EDUC., LEGAL RIGHTS OF CHILDREN 171 (1880), *reprinted in* THE LEGAL RIGHTS OF CHILDREN 31 (Sanford N. Katz et al. eds., 1974) (compulsory education laws "belong to the class of laws which are intended for the suppression of vice").

52. Teitelbaum & Harris, *supra* note 50, at 24-25.

of autonomy as they move through adolescence.[53] A child who fails to assert that autonomy may be described pejoratively as "tied to the apron strings." Modern psychological theory as well emphasizes the importance of separation, and a considerable body of current literature suggests that failure to achieve mature independence from parents produces a number of well-recognized psychological problems in adult children.[54]

3. Reciprocity as a Basis for Filial Support

A rule of family responsibility might rest on an ethical theory of reciprocity—that children owe support to parents because their parents supported them. Although this approach has found some judicial approval,[55] it is nonetheless difficult to justify in a systematic way. The duty of parents to children can be explained easily enough, because it may be said to arise from the parents' own adult decisions. A duty of children to their parents cannot be explained in this way. While children no doubt receive considerable benefit from their parents, we do not generally say that one who receives a benefit is obliged to return the good or its equivalent. There is, to be sure, a special quality to the rendition of the benefit by parents, because it flows from unconditional love. That quality suggests, however, that the benefit was given without an expectation of repayment rather than as the reflection of a culturally-based obligation.

The lack of such an expectation is illustrated by attitudes concerning the provision of medical care to elderly parents. Parents may think they should receive such care. They often believe that they are owed those services because of their social security payments over the years. Indeed, a survey of parents and their adult children found that most members of *both* groups stated that social security should provide needed income for the retired elderly.[56] Similarly, parents may think it right for private

53. *See* Al Katz & Lee E. Teitelbaum, *PINS Jurisdiction, The Vagueness Doctrine, and the Rule of Law*, 53 IND. L.J. 1, 17-23 (1977-78).

54. JOHNSON, *supra* note 44, at 117.

55. *E.g.*, Swoap v. Superior Court, 516 P.2d 840, 852 (Cal. 1973) (holding that duty of children to support needy parents was rational basis for classification of those who are required by law to reimburse state for aid granted to aged parents).

56. ALVIN SCHORR, U.S. DEP'T OF HEALTH AND HUMAN SERV., PUB. NO. 3-11953, ".

insurers to pay for medical care because of the premiums they themselves have paid to purchase that care. However, only ten percent of the elderly in the survey reported above believed that adult children should provide the needed support.[57]

If, as seems to be the case, neither parents nor their children feel a strong sense of reciprocal financial obligation, the explanation may be that such an obligation runs contrary to another cultural expectation: that children devote their resources to establishing their own lives and families. Jenny Baxter, a seventy-five-year-old Californian on old age security benefits, may have spoken for many in saying:

> When the child reaches maturity, he starts a new separate unit and in turn makes his contribution to life and society as did his parents . . . . The children should not be saddled with unjust demands that keep them at or near poverty level with no hope to escape it, just because a parent still breathes. And aged parents should not have to live their remaining lives facing the heartbreaking experience of being such a burden to their children. Many would prefer death but are afraid of retribution for taking their own lives.[58]

There are difficulties with a culturally based theory of reciprocity even if we assume that the parent wishes support or services from some source. Reciprocity is even more difficult to justify when it is applied to the situation where the disabled family member does not want care given at all. *Cruzan v. Director, Missouri Department of Health*[59] reflects just such a situation. The parents of an incompetent, comatose patient sought to terminate artificial feeding and hydration—a decision that the hospital refused to accept. The United States Supreme Court held

---

. . THY FATHER & THY MOTHER . . .": A SECOND LOOK AT FILIAL RESPONSIBILITY AND FAMILY POLICY 12 (1980) (citing LOUIS HARRIS & ASSOCS., NATIONAL COUNCIL ON AGING, THE MYTH AND REALITY OF AGING IN AMERICA (1975)). This attitude is not surprising. Social security is routinely described and justified as an insurance program and its benefits as "entitlements." While the government may not have managed or administered social security funds in a way that is consistent with these expectations, those who have contributed to this plan often believe they are owed care under its auspices.

57. *Id.* About one-third of the respondents felt that the elderly should provide for themselves. *Id.*

58. Swoap v. Superior Court, 516 P.2d 840, 864 (1973) (Tobriner, J., dissenting).

59. 110 S. Ct. 2841 (1990).

that while adults have a constitutionally protected liberty interest in avoiding unwanted treatment,[60] the State could appropriately require a high degree of certainty (here, "clear and convincing evidence") that the patient did not want treatment before accepting the view of a surrogate decision maker (here, a family member).[61] Of course, the critical consequence of a heightened burden of proof is that patients will be maintained even when it is more likely that they would prefer not to continue treatment and when their families accept that decision.[62]

It is hard to understand how notions of reciprocity, or any familial values, are served where the best estimate is that the parent wishes to avoid costly life-maintaining treatment and where provision of that treatment by the child means loss of economic resources necessary for other family opportunities.

4. Pragmatic Justifications for a Family Support Obligation

Finally, family responsibility laws might be justified on entirely pragmatic grounds; that is, such an obligation is necessary to avoid misuse of public resources. Family responsibility may be supported, for example, as a strategy for cost containment. It is widely said that, without a sense of personal cost, any resource, including health care, will be overused. To control this risk, both public and private health care systems typically require a co-payment as a device to reduce the unnecessary use of services.[63]

Placing responsibility for the care of parents on their children provides a similar vehicle for cost containment. Parents might not as readily seek services they do not need, or consume more of those services than necessary, if their children must pay for them. Conversely, children may supervise their parents' resort to health care if they are liable for its cost. Indeed, children may also play a broader caretaking role by encouraging parents to conserve their

---

60. *Id.* at 2851.

61. *Id.* at 2852-53. For a thoughtful critique of *Cruzan*, see Martha Minow, *The Role of Families in Medical Decisions*, 1991 UTAH L. REV. 1.

62. On the general subject of continuing treatment of critically and chronically ill persons who are incompetent at the time of decision, see JOHN A. ROBERTSON, THE RIGHTS OF THE CRITICALLY ILL 49-70 (1983).

63. Experiments requiring co-payments by groups of Medicaid recipients were also conducted precisely to determine their effect on medical care usage. *See* California Welfare Rights Org. v. Richardson, 348 F. Supp. 491 (N.D. Cal. 1972).

resources if the children face secondary responsibility for their parents' health care costs.

This argument has genuine appeal. However, it goes too far, even on its own terms. Family responsibility laws create a potentially unlimited obligation, which is far greater than is required for the limited purpose of reducing overuse of medical services. For that purpose, the obligation of family members would not be formulated as a general duty of support, but rather limited to an amount that would function as a co-payment, resembling the recipient's own responsibility under private insurance and Medicare.

A second form of misuse arises in connection with the Medicaid program: transfers from parents to children that are intended to impoverish the parents and therefore require the public to pay for costs that the parents could otherwise themselves have borne. As the rules now stand, it is quite possible for parents to give all their wealth to their children and quitclaim the family home as well, if this is done at least thirty months before the parent needs nursing home or home health care.[64] Medicaid will then be required to pay all nursing home bills and the adult children will retain what may be very substantial assets.

Now, this is surely an undesirable practice. Avoiding that practice does not, however, justify a general theory of filial responsibility for parental necessaries. For one thing, wealth transfers to create public responsibility are only available where the family members have strong ties to each other, the parents have some accumulated wealth which they are willing to pass on, post-transfer parental income is insufficient to pay for medical services, and the need for Medicaid eligibility is both apparent and predictable in time. Given these limitations, a general rule of family responsibility is far broader than the problem giving rise to the rule.

---

64. 42 U.S.C.A. § 1396p(c)(1) (West 1992) requires that all plans must provide for a period of ineligibility for nursing facility and home health services for persons who, within the 30 month period immediately before the date of institutionalization or application for assistance, dispose of resources for less than fair market value. It also provides that an individual will not be ineligible if the resources were a home transferred to a spouse, to a child under age 21 or disabled, or to a child living in the home for at least two years immediately before the date of institutionalization who provided care, thereby avoiding earlier institutionalization. *Id.* § 1396p(c)(2).

More limited solutions to this dubious practice are also available. The period of ineligibility may be extended further back in time: perhaps, to sixty months. Alternately, or additionally, children or other relatives might be made responsible for paying nursing costs to the extent they received assets from the donor for some period prior to the donor's eligibility for Medicaid, but not beyond the amount received.

IV. THE STRUCTURE OF LEGAL RULES AND THE FAMILY

Even if we were to identify a sufficient basis for imposing an obligation on children to support their needy parents, the desirability of doing so by law remains to be considered. Creation of a legal obligation involves not only the desirability of a rule but of a particular kind of rule to deal with family decisions.

A. The Tradition of Indirect Regulation

Laws expressing publicly-enforced guarantees and obligations take, or at least aspire to, a special form: proscriptive or prescriptive rules, stated in advance, of universal application. The content of a family responsibility law might be something like this: All children shall be responsible for supplying necessary goods and services to their parents when their parents are unable to do so themselves.

This rule seeks generality and clarity. Clarity, under current theories of justice, is required to assure that those subject to rules know and can plan for their obligations.[65] Both clarity and universality are required to assure that the rule antedates its imposition and that public force is not used because of illegitimate personal or social characteristics of persons before the court.[66] The question, then, is whether it is desirable to state such rules of obligation in the family context, creating a predictable basis for reliance both by family members and by society at large.

---

65. *See, e.g.*, Katz & Teitelbaum, *supra* note 53, at 4-5 (noting that statutory definiteness is designed to assure persons will be free from coercive intervention unless conduct has been previously proscribed).

66. *Id.* at 4-6.

The simplest answer may seem to be, "Well, why not?" There surely is ample normative support in cultural sentiment[67] and, for that matter, judicial pronouncements for such requirements. Courts speak routinely of a "family unit," a phrase that may imply legal responsibility among its members, and courts equally often describe the family as the basic foundation of civil society.[68]

Translation of even clear moral obligations into legal rules does not always follow, however, and the link between moral discourse and legal regulation is especially complex in family law. It is not that we lack a strong sense of how families should behave. Nor do we lack a sense of the importance of the modern family, sometimes bordering on the obsessive.[69] However, despite this consensus, what we do not find is a body of law converting these sentiments into detailed rules for the conduct of intact families: that is, families not disrupted by divorce. On the contrary, direct regulation of the family by specific and universal rules has long been thought inconsistent with respect for the family and indeed injurious to family relations.[70] Courts routinely declare that they will not regulate relations within the family,

---

67. *See supra* text accompanying notes 47-49 (discussing biblical and literary passages that demand/encourage loyalty to parents).

68. *E.g.*, Maynard v. Hill, 125 U.S. 190, 211 (1888); Reynolds v. United States, 98 U.S. 145, 165 (1878).

69. Nineteenth and twentieth century views celebrated the family as a haven from the jungle of economic and political life, and as the repository for nobler instincts that were foreign to the public arena. *See, e.g*, ROBERT L. GRISWOLD, FAMILY AND DIVORCE IN CALIFORNIA, 1850-1980, at 129-30 (1982); Philippe Aries, *The Family and the City in the Old World and the New, in* CHANGING IMAGES OF THE FAMILY 29 (Virginia Tufte & Barbara M. Myerhoff eds., 1979); John Demos, *Images of the American Family, Then and Now, in* CHANGING IMAGES OF THE FAMILY 43, 49-55 (Virginia Tufte & Virginia M. Myerhoff eds., 1979).

70. For various reasons, the technique of indirect regulation has long been the practice in family law. Early common law was not much concerned with the family as a social institution, largely limiting its attention to the individual who held or would hold tenure. Rules of inheritance, for example, were devised to assure that lands held in military tenure remained with a single person whose duties could easily be identified and enforced. Ultimately, this concern resulted in the institution of primogeniture, limiting inheritance of certain lands to the eldest surviving son. Other children, however important their relations within the family might have been, went largely without formal provision or recognition. Thus, for example, the entitlement of all children to inherit, recognized by canon law, found little acceptance in England and disappeared for the most part by the end of the fourteenth century. R.H. Helmholz, Legitim *in English History, in* R.H. HELMHOLZ, CANON LAW AND THE LAW OF ENGLAND 247, 253-258 (1987). Because the family as a social enterprise had little importance in the feudal scheme, there were few specific rules of family responsibility, except to avoid imposition on local communities.

often on the theory that intervention will adversely affect those relations. A Rhode Island court, for example, denied recovery to a minor who sued his parent and an insurance company for injuries arising from an automobile accident, on the following ground:

> Immemorially the family has been an important element of our civil society, one of the supports upon which our civilization has developed. Save as modified by the Legislature, in domestic affairs the family has remained in law a self-governing entity, under the discipline and direction of the father as its head. As part of the family order or arrangement are the related obligations and rights of the father and his minor child while the child remains in his household unemancipated. . . . Any proceeding tending to bring discord into the family and disorganize its government may well be regarded as contrary to the common law . . . .[71]

While the doctrine of parental immunity from suit by an unemancipated child has now largely been abandoned, that change has not resulted from a lessened concern for the family, but rather from the belief that such suits do not impair family harmony.[72]

Cases like *Balfour v. Balfour*[73] and *McGuire v. McGuire*[74] also come almost unbidden to mind, announcing that the "King's writ"—that is, law in its usual proscriptive and prescriptive form—does not run to the spousal relationship. This principle could be repeated almost endlessly in other domestic relations settings. While statutes and courts announce a parent-child support obligation, they rarely—apart from divorce—give it detailed meaning. There is such a thing as nonsupport, plainly, but its positive meaning is close to literal abandonment. The law does not require parents to leave their wealth to their children, continuing the practice of allowing parents to choose their

---

71. Matarese v. Matarese, 131 A. 198, 199 (R.I. 1925).

72. *See, e.g.*, Rousey v. Rousey, 528 A.2d 416, 420 (D.C. 1987) (availability of insurance reduces possibility of disrupting domestic tranquility).

73. [1919] 2 K.B. 571.

74. 59 N.W.2d 336, 342 (Neb. 1953) ("The living standards of a family are a matter of concern to the household, and not for the courts to determine . . . . As long as the home is maintained and the parties are living as husband and wife it may be said that the husband is legally supporting his wife and the purpose of the marriage relation is being carried out.").

testamentary dispositions.[75] And apart from family responsibility statutes, no legal rules specify the relations between adult children and their parents.

I do not mean to say that law does not affect the family and its relations. Indeed, the last section of this Article explores the profound importance of legal structures on the choices available to families.[76] However, the regulations that so heavily affect families are not part of any domestic relations code, but are found rather in health codes, internal revenue codes, and other bodies not identified as "family law." With respect to rules directly addressing families, legislatures and courts have generally preferred a scheme that localizes decisions rather than defines them in universal terms.

B. Rules and Family Responsibility

But why, it is reasonable to ask, do courts and legislatures prefer indirect regulation of family behavior, and what does this tradition tell us about family responsibility laws?

1. The Special Nature of Family Relations

Some of the reasons are, of course, historical.[77] Currently, however, they reflect a sense that certain aspects of modern legal theory do not operate well within the family context. A primary thrust of legal development since the eighteenth century has been the separation of legal personality from other social structures and relationships. This development reflected the need in a commercial and industrial society for predictability and, in turn, simplicity. In a complex modern economic and social system, legally reliable and calculable predictions are thought important for those who would engage in highly varied, long-term, and wide-ranging transactions. Moreover, those predictions must be abstractly calculable, so that

---

75. *See generally* Deborah H. Batt, *I Didn't Ask to Be Born: The American Law of Disinheritance and a Proposal for Change to a System of Protected Inheritance*, 41 HASTINGS L.J. 1197 (1990).

76. *See generally* Lee E. Teitelbaum, *Family History and Family Law*, 1985 WIS. L. REV. 1135 (discussing effect of judicial abstention and facilitative rules on how choices are made within family).

77. *See supra* note 70 (discussing historical reasons for indirect regulation).

individuals can make business judgments without the knowledge—no longer available—that comes with long-established relations or intimate knowledge of those with whom they dealt and the conventions they accepted.[78] It was, therefore, necessary to suppose that all persons had the same capacities and understandings.

These requirements of simplicity and predictability were ultimately supplied by law itself.[79] Contractual and delictual capacity in a world of strangers were founded on an assumption that all adult persons are equal for purposes of commercial transactions or tort liability. They are, consequently, neither rich nor poor, experienced nor naive, sophisticated nor foolish. The content of their transactions was also defined objectively by resort to trade practice in matters of contract and the reasonable person in torts.

By the same token, the legal orientation of modern society emphasizes the accomplishment of specific goals, generally disregarding elements unrelated to those goals, such as effects on nonparties or even effects on a party's life plans, if those plans are unrelated to the particular transaction. It is, for example, immaterial in a suit to enforce a contract that the defendant will be left unable to send her children to college.

We may suppose, although it is controversial, that the abstractness and goal-orientation of legal analyses, allowing specific calculation of outcomes, may be appropriate for commercial and tort matters.[80] Their applicability to domestic relations seems, however, less obvious. For one thing, family relationships—unlike most commercial and tort relationships—are conceived in lifelong and highly mutable terms. The variability and continuing quality of those relations make it difficult to accept

78. *See* NIKLAS LUHMANN, A SOCIOLOGICAL THEORY OF LAW 13-15 (Martin Albrow ed. & Elizabeth King & Martin Albrow trans., 1985).

79. *Id.* at 15.

80. Although much of the recent history of contract, for example, reflects efforts to create a reasonable balance between the reality of inequalities of power and the ideal of social equality. Among the familiar examples of legislative efforts in this direction are labor law, social security law, food and drug laws, antitrust laws, and laws regarding medical care. Judicial modifications include the doctrines of reliance and unconscionability. *See* GRANT GILMORE, THE DEATH OF CONTRACT (1974); Manfred Rehbinder, *Status, Contract, and the Welfare State*, 23 STAN. L. REV. 941 (1971).

the confined focus associated with commercial and delictual settings involving single transactions.

Moreover, unlike the organizations and institutions of modern society, families incorporate and are expected to incorporate diffuse, particularistic, and collective values and relations.[81] They are economic systems, as we have seen,[82] but they are simultaneously systems of social control, education, morality, and affective relations.[83] While specialized social institutions such as business entities and public agencies are designed to serve specific functions and determine their actions by reference to those purposes in a more or less formally rational way, the family accommodates all of its functions in each decision. Domestic economic decisions, unlike those of a corporation, are not influenced wholly, or often even largely, by economic rationality. What parents expect of children is heavily influenced by their own moral and political values and what children dedicate to parents is influenced in the same way.

2. *Differences Between Child Support and Parent Support*

It may seem that differences in the regulation of families and of other social groups, even if true in general, tell us nothing about family support obligations. We know, after all, that child-support obligations are defined and enforced every day in every state. If obligations from parents to children can be articulated, so can duties from children to parents and, of course, many states have done so.[84]

---

81. On family goals, see S.N. EISENSTADT, FROM GENERATION TO GENERATION: AGE GROUPS AND SOCIAL STRUCTURE 161 (1956).

82. *See supra* text accompanying notes 19-28 (discussing family's economic functions).

83. *See* Teitelbaum, *supra* note 18, at 813-18.

84. The California Civil Code, for example, provides that "[i]t is the duty of the father, the mother, and the children of any person in need who is unable to maintain himself by work, to maintain such person to the extent of their ability. . . . A person who is receiving aid to the aged shall be deemed to be a person in need who is unable to maintain himself by work." CAL. CIV. CODE § 206 (Deering 1990).

Section 12100 of the CALIF. WELF. & INST. CODE specifically authorized the state to proceed against an adult child who failed to contribute to the support of a parent. After a series of decisions upholding actions against children on the basis of Section 12100, the California Legislature repealed that provision in a statute, stating that no relative could be made responsible for the support of a recipient of medical or welfare aid. *See* Indest, *supra* note 9, at 236.

The obvious comparison with ordinary support provisions is, however, misleading. Within a going family, not affected by divorce or separation, the support obligation is expressed only very generally. Parental obligation to their children is usually formulated in terms of an obligation to supply "necessaries."[85] However, while what is "necessary" is defined to some extent objectively (according to "station in life"[86]) considerations of family privacy generally disincline courts to supervise the economic and social relations of parents and children when the family has not already been disrupted by divorce.[87] Therefore, the functioning nuclear group rather than legal rules chooses the values governing the kinds and extent of support.

Support of parents generally occurs in a very different setting. Relations between parent and child have been changed upon the latter's departure from the home, and in most cases they do not share a common residence. Separate residence remains the custom even where the parents are aged. Moreover, the consistent trend is toward less rather than more co-residence even for the elderly.[88] We have already seen that these residences are not linked in any direct economic fashion (through, for example, joint production or consumption of wealth). They are, as well, not tied in any other way that would supply internally generated understandings and expectations.

Of course, emotional relations among members of these separate households often endure. However, their strength and expressions are more various than in the nuclear home with its relatively clear sense of expectation and obligation from parents to children and children to parents. Between generations, prior to emancipation, obligation is easy enough to recognize, at least in general terms. Emancipation means, however, the end of that formal obligation in both directions. The content of a support obligation for family members who do not share a residence and

85. *E.g.*, Greenspan v. Slate, 97 A.2d 390, 392 (N.J. 1953) (parent's obligation to support child determined by parent's ability to pay, and child's fortune).

86. *E.g.*, Sharpe Furniture, Inc. v. Buckstaff, 299 N.W.2d 219, 224 (Wis. 1980) (purchase of sofa by wife considered legally "necessary" item).

87. *Cf.* Roe v. Doe, 272 N.E.2d 567, 570 (1971) (court would not require parent to pay for college education of daughter where daughter disobeyed father's instructions concerning living arrangements).

88. SCHORR, *supra* note 56, at 13.

do not maintain the roles associated with an intact family will—unlike child support in an intact home—largely be supplied by external sources rather than by the family itself.

The character of a family responsibility law, then, differs from the usual child support relationship in its lack of internally generated norms and limits. It also differs from the special situation where courts are prepared to supply the content of the support obligation; that is, when child support is ordered upon divorce.

In the event of divorce, legal rules seek to continue the level of support the children would have received had the marriage succeeded.[89] Because parents expect to care for their children, a rule expressing that calculation fits the paradigm of legal rules. Parents expect to arrange for their children's health care, education, and support and often do so from birth or even before. These anticipated costs become part of a complex calculus influencing other family wealth distributions. Parents often choose where they will work and live according to educational opportunities; insurance may have been purchased with the same goal; opportunities for travel or for the purchase of homes may be consciously deferred. To provide children necessaries according to family circumstances and hopes is an express part of parental expectations.

The capacity to calculate assumed by modern legal relationships is far weaker in the case of parental support than is true even for relations within the nuclear family. Financial assistance to parents is still typically unexpected and indeed often cannot be planned. Many children do not know the extent of their parents' public and private insurance protection and inquiries may not seem—or be—welcome. Even with such knowledge, adequate planning may not be possible. Private insurance policies selected by employers typically do not cover adult relatives other than spouses, or children for a few years of their majority. Employment and residence decisions are made, and appropriately made, without consideration of parental circumstances. Legal strategies that suppose an obligation of forethought and calculation, there-

89. *E.g.*, Smith v. Smith, 626 P.2d 342, 344 (Or. 1981) (recognizing that this often cannot be done).

fore, seem inappropriate to a situation in which planning is rarely done and often impossible.

V. ON SHARING RESPONSIBILITY

The discussion to this point has sketched various difficulties in the direct regulation of family relations in general and the adoption of family responsibility laws in particular. This is not, however, an argument against family responsibility or against recognizing a legal role in this respect. The discussion suggests, rather, that simple strategies of identifying and allocating responsibility—here, by substituting family for more general social responsibility—are undesirable.

A. The Complex Nature of Responsibility

One of the least satisfactory aspects of an approach that substitutes private (family) responsibility for public responsibility is its reliance on a categorical notion of responsibility itself. That approach seems to assume that responsibility exists or does not; that it rests in one place or another; that it is either borne or not borne.[90] Thus, the California Supreme Court, in upholding a family responsibility statute, observed that "a long tradition of law, not to mention a measureless history of societal customs, has singled out adult children to bear the burden of supporting their poor parents. This duty existed prior to, and independent of, any duties arising out of state assistance to the aged."[91]

A similar perception is implicit in discussions of abandonment of parents in hospitals; a practice sensitively described as "granny dumping" by the media. Both terms—abandonment and granny dumping—suggest a failure of moral and legal responsibility by children. This seemingly obvious conclusion leaps to mind, however, only because responsibility is understood as resting in one place only—here, with adult children.

---

90. *See* Katherine Bartlett, Dilemmas of Defining Intergenerational Responsibilities 3-5 (unpublished paper delivered at joint meeting of Family Law and Elderly sections of Association of American Law Schools, Annual Meeting of Association of American Law Schools, January, 1992) (on file with the *Utah Law Review*).

91. Swoap v. Superior Court, 516 P.2d 840, 849 (Cal. 1973).

Reality is, however, more complex than this notion of responsibility allows. If we consider the behavior involved in granny dumping, a person in need of care has been taken to an institution that exists for the purpose of giving care. Had the parent insurance or the child sufficient wealth, the behavior would be called an admission rather than an abandonment. To describe this conduct as abandonment must suppose the absence of a responsibility for the hospital to provide medical care for the parent who has been brought there. That supposition, in turn, goes unnoticed only because we have already located responsibility with the adult child.

I have suggested elsewhere[92] that the family might be imagined as the hub of a wheel, whose spokes connect the family to a variety of more specialized social institutions with which they share authority and responsibility. We have already seen that families distribute educational opportunity; however, public agencies are responsible as well for the availability, quality, and range of educational opportunity for children. Families function as agencies of social control insofar as they define whether certain conduct is deviant within the home, and decide how and when to enforce rules.[93] Police, juvenile courts, probation offices, and mental health providers also participate in the social control function, often together with the parents.

Shared rather than sole responsibility is, therefore, an important aspect of life even in a nuclear family. Moreover, allocations of responsibility to various sources are formally recognized. Looking only at the educational arena, parents are authorized and indeed expected to enforce their children's rights to programs that schools are required to provide (such as appropriate educational plans for disabled students).[94] At the same time, public agencies may enforce a parental obligation to assure school attendance by their children.[95]

---

92. *See* Teitelbaum, *supra* note 18, at 818.

93. *Id.* at 815-16.

94. *See, e.g.*, Board of Educ. v. Rowley, 458 U.S. 176, 182-83, 208-09 (1982) (parents have authority to enforce handicapped child's rights under government program).

95. The vehicle for enforcement of the state interest may be a criminal prosecution or a civil proceeding for neglect. The constitutionality of enforcing compulsory school laws against parents is taken for granted. *See, e.g*, State v. Kasuboski, 275 N.W.2d 101, 105 (Wis. Ct. App. 1978) (compulsory education reasonable unless parents' objection to education is "rooted in religious beliefs").

B. Shared Responsibility for Disabled Parents

We have already seen, in a general way, that families provide a great deal of support for their members,[96] quite apart from any prescriptive rules requiring that they do so. In 1988, approximately 3,602,000 persons provided financial support to more than 5,400,000 adults not living in the household with them.[97] Almost all of those receiving such support were relatives[98] and parents made up the largest group of non-household adults receiving support.[99] These numbers, not surprisingly, had increased even since 1986. The number of children receiving support decreased in that brief period both absolutely (from about 7.9 million to about 7 million) and as a proportion of all persons receiving support (from 65.0% to 56.2%). Concomitantly, the number of adults receiving support rose by more than twenty percent and the number of parents increased by almost thirty percent (to 1,665,000).[100] The average annual payment by children to their parents was $1330.[101]

Family members support their relatives, including their parents, because they have a special sense of loyalty to each other. At the same time, their decisions to provide support are not independent of other social agencies, nor are they entirely the expressions of actual family preferences. The capacity of families to make various decisions and the ways in which they express

---

96. *See supra* text accompanying notes 26-28.

97. BUREAU OF THE CENSUS, U.S. DEP'T OF COMMERCE, WHO'S HELPING OUT? SUPPORT NETWORKS AMONG AMERICAN FAMILIES: 1988, at 2 (1992) [hereinafter WHO'S HELPING OUT?].

98. Only 2% of the recipients were known to be non-relatives; the relationship of an additional 3.7% was not determined. *Id.*

99. *Id.* Of the adult recipients 1,665,000 of the 5,429,000 (31%) were parents. The next largest group was adult children (approximately 1.2 million). Only slightly over one-half million recipients were ex-spouses. *Id.*

100. *Id.*

101. *Id.* at 6-7. While sons are more likely than daughters to provide financial assistance to their parents, daughters are more likely to provide other kinds of help. *Id.* at 6, 8; Robyn Stone et al., *Caregivers of the Frail Elderly: A National Profile*, 27 GERONTOLOGIST 616, 617 (1987).

It should be added that the trend to increased support of parents has been with us at least since 1960 and will likely increase sharply until the Baby Boom generation reaches 65 in 2030. At that time, the ratio of retirement-aged persons to the working-age population will have increased from its current level of 20 per 100 persons to 37 per 100. WHO'S HELPING OUT?, *supra* note 97, at 1.

their sense of special obligation are affected by decisions made by other institutions.

Provision of nursing care for an elderly, partially disabled parent may illustrate the interrelationship of private and public activities. The provision of such care cannot be regarded as "private" in the descriptive sense that parents would prefer that their children, rather than others, provide such support or that children prefer providing support to having it come from others. In fact, many parents would actually prefer that others provide needed care and believe that those others, whether private or public, are or should be obliged to do so. There is also considerable evidence that many people do not wish to provide personal care for their elderly relatives.[102]

Therefore, if children bear the support obligation they do so not as a natural preference, but because parents do not have access to care through private or public sources. That lack of access, in turn, is the result of substantial and notorious limitations on the extent and coverage of both private and public insurance and benefits schemes.

Medicare, a federal program available to everyone sixty-five or older who is eligible for social security benefits (and to certain others on social security disability) is regarded as an insurance program covering some hospital and medical costs. However, its coverage does not reach certain substantial costs associated with disability and only partially covers other costs. Of particular importance for present purposes is that Medicare covers only a part of the cost of skilled nursing care and none of the cost of custodial or nonskilled care.[103]

---

102. *See infra* text accompanying note 112 (discussing difficulties of caring for disabled relatives).

103. In addition to responsibility for paying a deductible amount ($596 in 1990), Medicare requires a substantial "co-insurance" payment ($149 per day) for each day of hospitalization after 60 days. MATTHEWS, *supra* note 26, at 6:18. If long-term nursing care is required, coverage is limited or non-existent. Skilled nursing care is partly compensated; however, the patient is personally responsible for all costs after 100 days and for $65 per day for days 20-100. *Id.* at 6:20. Medicare pays for all covered costs only for the first 20 days. *Id.*

Nonskilled or custodial nursing care is not covered at all by Medicare. Moreover, Medicare does not pay for full-time nursing care at home, for housekeeping services, or for drugs or other therapies administered in the home. *Id.* at 6:20. Medicare Supplemental Medical Insurance pays for physicians' services and related expenses. It is, however, subject to an annual deductible and ordinarily pays only 80% of "reasonable" charges for

Supplementation in some degree is available through Medicaid. Medicaid reaches both unskilled care and the skilled nursing care costs omitted by Medicare.[104] Unlike Medicare, however, Medicaid is a welfare program administered by the states.[105] Nobody is qualified whose assets and income do not fall below levels set by the states, pursuant to broad federal guidelines. Those levels are in fact very low. Medicaid currently covers only slightly more than one-half of those whose income falls below 115% of the federal poverty standard. Both the proportions and absolute number of those eligible for Medicaid have decreased over the last decade.[106]

Even when Medicaid is available, its resources carry a heavy price. If the beneficiary is the only income producer, eligibility for Medicaid benefits requires not only the recipient's impoverishment but, in effect, that of her spouse. The "community" (non-institutionalized) spouse is entitled to an allowance of 150% of the official poverty line, not to exceed $1500 per month.[107] Consequently, family members may feel obliged to pay the costs of institutionalization, even when Medicaid would be available, to avoid reducing the non-recipient spouse (who is, after all, also their parent) to near poverty-level subsistence.

Moreover, reliance on Medicaid means that providing care for family members must become, in many cases, an all-or-nothing proposition. Family members cannot do "a little," or "as much as

---

those services.

104. Studies of nursing home expenditures indicate that about one-half of nursing home residents have Medicaid as the primary source of payment. Atrostic & Conly, *supra* note 13, at 181. Private insurance and Medicare have traditionally paid for only a small fraction of all nursing home costs. *Id.*

105. The Medicaid Program is authorized under Title XIX of the Social Security Act, codified at 42 U.S.C. § 1396 (1988). The initial decision by a state to participate is voluntary, but once a state participates, it must comply with federal requirements. Within those requirements, states implement their own programs.

106. BARRY R. FURROW ET AL., THE LAW OF HEALTH CARE ORGANIZATION AND FINANCE 234 (1991).

107. The 150% allowance is effective as of July 1, 1992. 42 U.S.C.A. § 1396r-5(d)(1), (2) (West 1992). The community spouse is also entitled to an "excess shelter allowance" in some cases.

Until recently, the same result—impoverishment—also would have occurred even when the patient was not the income-producer. The Medicare Catastrophic Care Act of 1988 substantially mitigated the consequences of previous law requiring that the assets and income of the non-recipient spouse be deemed available to the recipient and applied against the services provided. The Act now provides for separate treatment of resources after eligibility is established. 42 U.S.C.A. § 1396r-5(c)(4) (West 1992).

they can," for a disabled relative who is receiving Medicaid benefits. Any income regularly received by the patient—including gifts from children—will be applied towards determining eligibility for Medicaid and, assuming eligibility, toward the cost of services provided.[108] Children cannot, therefore, feel that they are "sharing" support for needy relatives by providing something that makes their lives better or is even noticeable. These rules create an obvious disincentive for family members to contribute to the support of their relatives.[109] At the same time, a scheme that discourages contributions may also compound the guilt and frustration of adult children who can find no way to express in a tangible and continuing, but limited, fashion their sense of obligation to their parents.

When neither private nor public insurance or benefits is available, care will be provided—if at all—by family members themselves. This alternative, too, significantly affects the lives of family members and is dramatically affected by practices of other agencies.

One possibility is for children to purchase the needed care for their parents. This would be expensive. The average cost of institutional care was estimated at $17,500 per year in 1985[110] and is surely considerably higher now. The cost of purchasing care is aggravated by various tax regulations, which should be considered another spoke in the wheel of family regulation. Current income tax provisions do not allow children to deduct contributions for the support of their parents, unless the support is so extensive as to qualify the parent as a dependent.[111] Gift tax liability also exists if the amount contributed exceeds the annual exclusion, unless contributions can be made directly to a qualified charitable organization. Again, both relief and an opportunity to share are denied.

---

108. MATTHEWS, *supra* note 26, at 7:5.

109. *See* Judith G. McMullen, *Family Support of the Disabled: A Legislative Proposal to Create Incentives to Support Disabled Family Members*, 23 U. MICH. J.L. REF. 439, 450-52 (1990) (discussing support of disabled children under SSI regulations).

110. Atrostic & Conly, *supra* note 13, at 181.

111. An individual taxpayer may deduct medical expenses paid for himself or herself, a spouse, and dependents. I.R.C. § 213(a) (1988). A dependent is a relative, including a parent, who receives more than 50% of his or her support from the taxpayer. *Id.* § 152(a).

Nor is income tax relief available for the competing family expenditures most likely to be affected by the need to use resources for parental support. When, as is common enough, children providing support to their parents themselves have children of high school or college age and wish to send those children to private schools, colleges, or graduate programs, those expenses also must be met with after-tax dollars. Neither private educational expenses paid by parents in an intact family nor child support payments in the event of divorce is deductible. These simultaneous support expenses must be managed, if at all, without income tax recognition or assistance.

If nursing care cannot be purchased, it must be provided personally. Care of the disabled elderly is also costly, especially in terms of stress, emotion, and time. The conventional image of caregiving as an altruistic activity that is not only accepted but welcomed has been severely criticized.

> [T]here appears to be no recognition of the fact that many people, if given a true choice, would prefer to abstain from the caregiving role altogether. The more common picture that is painted is one of a loving spouse who tends to the needs of the disabled person and only gives up this role when personal resources are exhausted. The facts that the disabled person no longer functions as a companion, is demanding, and may even be abusive are rarely acknowledged. While caregiving may be a choice when the disabled person responds with love and actively participates in a mutually rewarding relationship, there is considerable reason to believe that for many people caregiving is a relationship of bondage from which a spouse is unable to escape. Perhaps the most compelling support of this scenario are the findings from a recent study . . . that show that the group of caregivers who had the least burden and the highest scores on a morale scale after 12 months of participation were those caregivers whose elderly relative died during that period. . . . Caregivers whose relative had moved to the nursing home faired [sic] better than did those who continued to care for the older person.[112]

112. Edgar F. Borgatta & Rhonda J.V. Montgomery, *Aging Policy and Societal Values*, *in* CRITICAL ISSUES IN AGING POLICY 7, 22 (Edgar F. Borgatta & Rhonda J.V. Montgomery eds., 1987) (citations omitted).

Moreover, the decision to provide personal care may prevent caregivers, usually women,[113] from working in the paid economy, creating further financial and personal costs for that family member. And both sets of costs are multiplied when the caregiver is a single parent.

C. Refocusing Family Responsibility

If we accept the interrelationship between the conduct of families and the policies and practices of intermediate social institutions, the problem of family responsibility becomes very complex indeed. At the same time, that interrelationship may provide a different focus for efforts to deal with the ethical and social complexity of family decision making.

To begin with, we should abandon examination of responsibility in simple terms. To locate that responsibility with the family rather than with the state, as prescriptive rules directed to relatives propose, takes too little account of the effects of other bodies of law on family life. Instead, approaches to family responsibility should focus on the practices of both public and private institutions whose activities influence family decisions and life.

Such a shift in focus entails some substantial change in our thinking about these social institutions. As things now stand, we consider their practices on their own terms. Institutional policy analysis looks in the first instance to relatively specific sets of purposes and concerns associated with that institution. Effects on families are marginally important at best. Federal health care policies, for example, are guided by a preference for an "insurance" approach that only grudgingly extends "welfare" to the very poor, and then only on condition that virtually all of the recipient's own income and resources are exhausted. Consequently, the effects of health care policies on families now remain very harsh, despite recent ameliorative efforts.[114] Health care providers, for their part, no doubt wish that costs were lower, but price decisions are heavily governed by their interests in "appropriate" compensation for staff, delivery of what they regard as good medical care, and

---

113. *See, e.g.*, Stone, *supra* note 101, at 620.

114. *See supra* note 107 (discussing mitigating effect of Medicare Catastrophic Care Act of 1988).

perhaps also by considerations of prestige associated with certain kinds of facilities, equipment, or services. Tax policy is governed by the need to generate income and by decisions to assist competing taxpayer groups who are far more politically visible than families.

The focus proposed here requires a different approach, which examines the effects on family choice of policies adopted by each, and also by all, of the agencies whose practices profoundly affect family life. Moreover, this reconsideration should directly address ethical and economic questions of the sorts already suggested, which are likely to arise only secondarily when the focus is on the functioning of institutions rather than on their effects on families.

Take, for example, the scope of care available to the elderly: How much medical care and what kind of care will we provide to the elderly? Placing responsibility for the care of the elderly on family members, as family responsibility laws undertake, tends to obscure or at least minimize the importance of this issue. If the duty belongs with the children of disabled parents, the easy conclusion is that we expect children to do their duty. We have, on this approach, no need to consider closely the extent and implications of that duty.

There are, however, distributive questions that should not be ignored. One of the implications of placing responsibility for the medical care of the elderly on family members is that the availability of medical care will depend greatly on family wealth rather than some other basis, such as degree of need or relative importance of the care provided. This proposition is, of course, highly controversial generally and with regard to health care specifically. That issue is far more likely to be examined carefully when our focus is on the responsibility of the health care system and the government, rather than on a general duty of family members.

Similarly, a narrow focus on family obligations masks issues of the provision of medical care. Whether certain services should be provided regardless of the probable desires of the patient and the known desires of the relatives is the most obvious of these questions. A related, but less obvious and sympathetic question, is whether all forms of treatment considered appropriate by law and medical practice are more important than, for example, providing a college education for one's children or reserving funds for one's own old age.

In effect, family responsibility laws *do* answer these questions by implication: the appropriateness of treatment is a medical question which, once decided, answers all of the other questions. These decisions are made *en passant*. But they are genuinely important and difficult questions which deserve careful consideration. Because a family responsibility rule is directed to one set of family members, however, the desirability of these and the myriad other distributive choices facing particular families will not receive direct consideration.

Our primary concern has been with decisions obviously falling within the concern of health care institutions. Refocusing must, however, recognize that choices by families are affected not only by other social institutions but by a number of such institutions. To take one example, we have already seen that the level of cost associated with care for disabled relatives is affected by the tax treatment of payments as well as by health care charges.[115] Accordingly, examination of the relationship between tax policy and family contributions to the care of relatives must be part of any inquiry that takes family responsibility seriously.

This examination might include ethical issues arising from decisions to permit resource distributions favoring members of one's own family, on the one hand, and by cultural values urging members of a family to care for their relatives, on the other. If, as an ethical matter, the elderly disabled should have equal access to medical care, tax advantages for family contributions to such care would not be sensible. If, on the other hand, such contributions should be approved or even encouraged, some tax advantage is not only defensible but desirable.

Examination of tax policy regarding medical care of family members might also consider the alternate uses to which families might put any tax advantage provided for care of elderly members, including educational expenses for other family members or purchase of insurance for their own seniority. We have already observed that care of parents is only one set of family-related expenses for adult children. Tax laws sensitive to family choices might, therefore, offset costs directed to one legitimate family

---

115. *See supra* text accompanying notes 110-11 (discussing tax treatment of health care payments).

purpose with deductions or credits for other important family purposes.

VI. CONCLUSION

This is not the place to examine all of the bodies of law that should be considered even in connection with the transfers described above. My concern is with the way in which responsibility is understood and expressed.

I have only meant to suggest that a candid and responsible approach to family support should accept as inevitable that this undertaking will be shared and seek to distribute aspects of that undertaking consciously, fairly, and with an appreciation of the likely effects of decisions in one domain on those in others. The issue of family obligation is a complex question of social policy requiring a far broader frame of reference than family responsibility law requires or even permits.

The Problem of Proxies with Interests of Their Own: Toward a Better Theory of Proxy Decisions*

John Hardwig**

A seventy-eight-year-old, married woman with progressive Alzheimer's disease was admitted to a local hospital with pneumonia and other medical problems. She recognized no one and had been incontinent for about a year. Despite aggressive treatment, the pneumonia failed to resolve and it seemed increasingly likely that this admission was to be for terminal care. The patient's husband (who had been taking care of her in their home) began requesting that the doctors be less aggressive in her treatment and, as the days wore on, he became more and more insistent that they scale back their aggressive care. The physicians were reluctant to do so, due to the small but real chance that the patient could survive to discharge. But her husband was her only remaining family, so he was the logical proxy decision maker. Multiple conferences ensued, and finally a conference with a social worker revealed that the husband had recently proposed marriage to the couple's housekeeper and she had accepted.

I.

Patient autonomy is the cornerstone of our medical ethics. Given this commitment to autonomy, proxy decisions will always strike us as problematic; it is always more difficult to ensure that the wishes of the patient are embodied in treatment decisions when someone else must speak for the patient. And proxy decisions are especially disturbing when we fear that the proxy's judgment is tainted by his own interests, so that the proxy is covertly requesting the treatment *he* wants the patient to have, rather than the treatment the *patient* would have wanted. This

---

** Professor of Philosophy and Humanities, East Tennessee State University; Professor of Medical Ethics, James H. Quillen College of Medicine, East Tennessee State University. I wish to thank Eric H. Loewy and especially Mary R. English for many helpful comments on this paper.

problem of interested proxies is exacerbated by the fact that we seek out proxies who often turn out to have strong interests in the treatment of the patient. We do this for two reasons: first, those who care deeply for the patient are more likely than others to really want the best for the patient; and second, those who are close to the patient are generally most knowledgeable about what the patient would have wanted. This familiarity allows us to apply the "substituted judgment" standard of proxy decision and, given a commitment to autonomy, substituted judgment is an ethically better basis for proxy decision making than the "reasonable person" or "best interest" standard.

The apparent alternative would be proxy decisions made by outsiders—physicians, court appointed guardians, or ethics committees. We must learn to recognize that such outsiders also have interests of their own and that their proxy decisions may also be influenced by these interests. But the more common worry about outsiders is that they rarely know the patient as well as members of the patient's family, and their concern about the individual patient does not run nearly as deep. Proxies who are members of the patient's family have a difficult time ignoring their own interests in treatment decisions precisely because they—unlike outsiders—are so intimately involved with the patient and have so much at stake.

Thus, it seems that our theory of proxy decisions has boxed us into a catch-22 situation. Knowledgeable about patient wishes usually means close, but close almost always means having interests of one's own in the case. Disinterested usually means distant, and distance usually brings with it less real concern, as well as lack of the intimate knowledge required to render a reliable substituted judgment.

I will urge that the reservations we have about interested family members and their proxy decisions are partly of our own making. The accepted theory of proxy decisions is deeply flawed and must be recast. Our medical practice is, I believe, often better than the conventional theories of proxy decision making. Nonetheless, some of our deepest worries about proxy decision makers grow out of the morally inappropriate instructions we give them.

If the current theory about proxy decisions for incompetent patients is mistaken, the accepted view of decisions by *competent*

patients will have to be modified as well. I will discuss competent patient decision making very briefly at the end of the paper.

II.

The husband in this case seems a perfect scoundrel. The physicians involved in the case all believed that he should be disqualified as a proxy decision maker, due to his obvious conflict of interest and his patent inability to ignore his own interests in making decisions about his wife's care. There was no reason to believe that the patient would have wanted to limit her treatment, so the conclusion seemed inescapable that the husband was not faithfully discharging his role as proxy decider.

Both traditional codes and contemporary theories of medical ethics hold that physicians are obligated to deliver treatment that reflects the wishes or the best interests of the patient,[1] and that the incompetence of the patient does nothing to alter this obligation.[2] There is similar unanimity about the responsibilities of a proxy decision maker. Proxy decision makers are to make the treatment decisions that most faithfully reflect the patient's wishes or, if those wishes cannot be known, the best interests of the patient. If the proxy does not do so, commentators almost uniformly recommend that physicians reject the proxy's requests and have recourse to an ethics committee or to the courts.

Despite the impressive consensus of both traditional codes and contemporary theories of medical ethics, I was intrigued by this case and pressed the attending physician for more details.

---

1. *See* Ludwig Edelstein, *The Hippocratic Oath: Text, Translation and Interpretation*, BULL. HIST. MED. 1, 3, 20-24 (Supp. I 1943); The World Medical Association, *Declaration of Geneva*, WORLD MED. J. 10, 12 (Supp. 1956); The World Medical Association, *International Code of Medical Ethics*, 1 WORLD MED. ASS'N BULL. 108, 111 (1949).

2. *See* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 73, 79-82 (3d ed. 1989); ALLEN E. BUCHANAN & DAN W. BROCK, DECIDING FOR OTHERS: THE ETHICS OF SURROGATE DECISION MAKING 10 (1989); JAMES F. CHILDRESS, WHO SHOULD DECIDE? PATERNALISM IN HEALTH CARE 102-107 (1982); THE HASTINGS CENTER, GUIDELINES ON THE TERMINATION OF LIFE-SUSTAINING TREATMENT AND THE CARE OF THE DYING 18-29 (1987) [hereinafter HASTINGS CENTER GUIDELINES]; EDMUND D. PELLEGRINO & DAVID C. THOMASMA, FOR THE PATIENT'S GOOD 3-6 (1988); 1 PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, MAKING HEALTH CARE DECISIONS: A REPORT ON THE ETHICAL AND LEGAL IMPLICATIONS OF INFORMED CONSENT IN THE PATIENT-PRACTITIONER RELATIONSHIP 2-6 (1982); ROBERT M. VEATCH, A THEORY OF MEDICAL ETHICS 22-26 (1981).

"Why is the husband in such a hurry? Does he hope that his wife will die? But she is dying anyway. Is he afraid that she might not die?" "No," the attending physician responded, "his worries are primarily financial. He is afraid that he'll lose his house and all his savings to medical bills before she dies. Since the housekeeper has no assets, they will then be left poverty-stricken."

To some, this seems even worse: not only has the husband allowed his own interests to override considerations of what is best for his wife, crass financial considerations are what concern him and he has let those kinds of considerations predominate. If his decision is not altogether self-centered, it is only because he is concerned about his fiancee's future as well as his own. But married men are not supposed to have fiancees.

I do not necessarily want to argue that the husband made the correct decision. And I do not know enough about him to judge his character. But I do think his decision should not be rejected out-of-hand, as patently inappropriate.

First, I do not think that we can simply assume that the presence of another woman means that he was insensitive to his wife's interests. I know couples who have divorced without losing the ability to genuinely care about each other and each other's interests. Second, while divorcing a long-standing wife simply because she is now demented is difficult—"How can I abandon her at a time when she is so vulnerable?"—remaining married to a woman with Alzheimer's—an increasingly unreachable, foreign wife—is difficult, too. His wife's dementia undoubtedly meant increasing isolation for him as well as for her. And given that, his search for companionship does not seem unreasonable or morally objectionable. Third, the husband also had been the patient's primary care giver for *years* without any prospect of relief or improvement. He probably longed for a chance to spend his few remaining years free of the burdens of such care. And, finally, supposing the husband to be an adherent of traditional values, he would not be able to simply "live with" the housekeeper, nor to consider himself no longer married while his wife was still alive, nor yet to accept medical care with no intention of trying to pay for it. Perhaps more "liberal" attitudes toward marriage and the payment of debts would have served his wife better. But perhaps not.

I have no doubt that the husband's proxy decisions were influenced by his own interests. Given the reasonableness and magnitude of the interests he had at stake, it is hard to see how he could ignore them. "How can *we* ignore his interests?" I wondered. "And how can we reasonably ask him to ignore them?" I do not think we can.

The attending physician and I did not get further on this case than my suggestion that the husband's concern about his financial future was an appropriate consideration in deciding on a course of treatment for the patient. The physician was shocked that I thought this kind of consideration was relevant.

Nevertheless, we limit treatment all the time in an effort to save money for the government or a health maintenance organization. We develop theories of rationing and "costworthy" medicine to justify such decisions.[3] We regularly deinstitutionalize people, partly to limit the cost of the care we as a society must provide. We limit the number of nursing home beds available for this man's wife and other Alzheimer's victims for the same reason. We thus force the burden of long-term care onto the families of the ill. And then we tell them that they must not consider their own burdens in making treatment decisions. I cannot make ethical sense of this.

We consider *our* pocketbooks, so how can we in good conscience tell proxies that they must ignore the much greater impact of aggressive treatment on their personal financial futures? I would insist that financial considerations for a seventy-five-year-old man with limited means are never trivial. We must recognize that for him, nothing less is at stake than the quality of the rest of his life, including, quite likely, the quality of his future health care.

If we find it morally repugnant that proxies decide to limit treatment due to the burdens of long-term care on the family, then

---

3. *See, e.g.*, DANIEL CALLAHAN, SETTING LIMITS: MEDICAL GOALS IN AN AGING SOCIETY 133-153 (1987); NORMAN DANIELS, JUST HEALTH CARE 1-18 (1985); Roger W. Evans, *Health Care Technology and the Inevitability of Resource Allocation and Rationing Decisions*, 249 JAMA 2208, 2208-11 (1983); E. Haavi Morreim, *Fiscal Scarcity and the Inevitability of Bedside Budget Balancing*, 149 ARCHIVES INTERNAL MED. 1012, 1012-1015 (1989); Lester C. Thurow, *Learning to Say "No,"* 311 NEW ENG. J. MED. 1569, 1569-72 (1984); Robert M. Veatch, *Justice and the Economics of Terminal Illness*, HASTINGS CENTER REP., Aug./Sept. 1988, at 34-36.

it is incumbent upon us to devise an alternative to our present system under which families deliver seventy-five percent of the long-term care. And until we have such an alternative in place, we dare not direct the husband that he must ignore the impact of treatment decisions on his own life. For *we* do not ignore the impact of such decisions on our lives. Moreover, the burdens of his wife's treatment to him may well outweigh any benefits we might be able to provide for her.

III.

There are, of course, many cases like this, in which optimal care for a patient will result in diminished quality of life for those close to the patient. This care can be a crushing financial burden, depriving other family members of many different goods and opportunities. But the burdens are by no means only financial; caring for an aging parent with decreasing mental capabilities or a severely retarded child with multiple medical problems can easily become the social and emotional center of a family's existence, draining away time and energy from all other facets of life. What are we to say about such cases?

I submit that we must acknowledge that many treatment decisions inevitably and dramatically affect the quality of more lives than one. This is true for a variety of very different reasons. First, people get emotionally involved with one another, and whatever affects those I love affects me too. Second, people live together, and important changes in one member of a living unit will usually have ramifications for all the others as well. Third, the family is still a financial unit in our culture, and treatment decisions often carry important financial implications that can radically limit the life plans of the rest of the family. Fourth, marriage and the family are also legal relationships, and one's legal status hinges on the life or death of other members of the family.

Finally, treatment decisions have an important impact on the lives of others because we are still, to some extent, loyal to one another. Most of us still do not believe that family and friendships are to be dissolved whenever their continued existence threatens one's quality of life. I know of a man who left his wife the day after she learned that she had cancer because living with a cancer-

stricken woman was no part of his vision of the good life. But most of us are unable or unwilling to disentangle ourselves and our lives from others as soon as continuing involvement threatens the quality of our own lives.

This loyalty is undoubtedly a good thing; without it we would have alliances for better but not for worse, in health but not in sickness, and until death appears on the horizon. It is a good thing even though it sometimes brings about one of the really poignant ironies of human existence; sometimes it is precisely this loyalty which gives rise to insoluble and very basic conflicts of interest, as measures to promote the quality of one life undermine the quality of others. If the husband in the case we have been considering had simply divorced his wife when she was diagnosed as having Alzheimer's, she would have died utterly alone. As such, only her own interests would have been relevant to her treatment. Her husband's loyalty—impure though it may have been—has undoubtedly made her life with Alzheimer's much better for her. But it also makes her treatment not simply her own.

Now, if medical treatment decisions will often dramatically affect the lives of more than one, I submit that we cannot morally disregard the impact of those decisions on all lives except the patient's. Nor can we justify making the interests of the patient predominant by claiming that medical interests should always take precedence over other interests. Life and health are important goods in the lives of almost everyone. Consequently, health-related considerations are often important enough to override the interests of family members in treatment decisions, but not always. Although persons become "patients" in medical settings and medical settings are organized around issues of life and health, we must bear in mind that even life or death is not always the most important consideration. We must beware the power of the medical context to subordinate all other interests to medical interests, for non-medical interests of non-patients sometimes morally ought to take precedence over medical interests of patients.

Because medical treatment decisions often deeply affect more lives than one, proxy decision makers must consider the ramifications of treatment decisions on all those who will be importantly affected, including themselves. Everyone with important interests at stake has a morally legitimate claim to consideration; no one's

interests can be ignored or left out of consideration. And this means nothing less than that the morally best treatment in many cases will not be the treatment that is best for the patient.

An exclusively patient-centered ethics must be abandoned. It must be abandoned, not only—as is now often acknowledged—because of scarce medical resources and society's limited ability to meet virtually unlimited demands for medical treatment. It must be abandoned, as well, because it is patently unfair to the families of patients. And if this is correct, the current theory of proxy decisions must be rejected in favor of an ethics that attempts to harmonize and balance the interests of friends and family whose lives will be deeply affected by the patient's treatment.[4]

IV.

There is a second, related point. Arguably, there is a presumption that substituted judgment is the morally appropriate standard for a proxy decision maker. But this can be no more than a *presumption* and it can be overridden whenever various treatment options will affect the lives of the patient's family. In fact, substituted judgment is the appropriate standard for proxy decision making in only two special (though not uncommon) situations: first, when the treatment decision will affect only the patient, or second, when the patient's judgment would have duly

4. There are a few scattered references which acknowledge that the interests of the patient's family may be considered. At one point, the President's Commission states that "[t]he impact of a decision on an incapacitated patient's loved ones may be taken into account." PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBS. IN MED. AND BIOMED. AND BEHAVIORAL RES., DECIDING TO FOREGO LIFE-SUSTAINING TREATMENT 135-36 (1983) [hereinafter PRESIDENT'S COMM'N REPORT]. The HASTINGS CENTER GUIDELINES, *supra* note 2, counsels consideration of the benefits and burdens to "the patient's family and concerned friends," but only in the special case of patients with irreversible loss of consciousness. *Id.* at 29. Buchanan and Brock devote one page of their impressive work, *Deciding for Others*, to the "limits on the burdens it is reasonable to expect family members to bear." BUCHANAN & BROCK, *supra* note 2, at 208. But these are only isolated passages in large, systematic works and they do not inform the over-all theory developed in these works. The discussion of neonatal care is the only place I know where the interests of members of the patient's family have received systematic attention. *See, e.g.*, Carson Strong, *The Neonatologist's Duty to Patients and Parents*, HASTINGS CENTER REP., Aug. 1984, at 10. The fact that many ethicists seem willing to consider family interests in the case of newborns but not in the case of older patients suggests that we may not really consider newborns to be full-fledged persons.

reflected the interests of others whose lives will be affected. In other situations, proxy deciders should make decisions that may be *at odds with* the known wishes of a formerly competent patient.

Consider again the case with which this paper began. I did not know the patient, and I have no idea what kind of a person she used to be. Let us, then, consider two, rather extreme hypotheses about her character. On one hand, suppose that the patient had been a very selfish, domineering woman who, throughout their marriage, had always been willing to sacrifice her husband's interests to her own. If so, we can reliably infer that she would now have ignored her husband's interests again, perhaps even ridden roughshod over them, if she could have gotten something she wanted by doing so. Therefore, we can conclude that she would have demanded all the medical treatment available, regardless of costs to him. We can even imagine that she would have relished her continuing power over him and her ability to continue to extract sacrifices from him. Obviously, her husband would know these facts about her. The substituted judgment standard of proxy decisions would have us conclude that if that is the kind of woman she was, this would *increase* her husband's obligation to make additional sacrifices of his interests to hers.

Suppose, on the other hand, that this woman had always been a generous, considerate, unselfish woman; deeply sensitive to the interests of her husband and always ready to put his needs before her own. If that is the kind of woman she was, the theory of substituted judgment allows—strictly speaking, even *obligates*—her husband to sacrifice her interests once again by now demanding minimal care for her. After all, he knows that is what she would have done, had she been competent to make the decision. Even if he wanted to give her the very best treatment as an expression of love or gratitude for her concern for him throughout their lives, substituted judgment would require that he ignore those desires. That is what *he* wants for her, not what *she* would have chosen for herself.

But surely that is exactly wrong. The theory of substituted judgment has it backwards. Loving, giving, generous people deserve to be generously cared for when they can no longer make decisions for themselves, even if they would not have been generous with themselves. And what do selfish, domineering, tyrannical people deserve? The answer to that question depends

on your ethical theory. Perhaps neglect, maybe even retribution, are justified or at least excusable. Perhaps tyrannical behavior releases the family from any *special* obligation to care for the now incompetent tyrant. But unless you believe that good people should not be rewarded for their virtues, you will agree that caring, giving individuals deserve better care than domineering, self-centered individuals.

Where did we go wrong? What led us to widespread acceptance of the theory of substituted judgment? The major mistake was the one we have been considering—the mistake of believing that medical treatment affects only the life of the patient, or that its impact on other lives should be ignored. If the patient's interests are the only ones that ought to shape treatment decisions, those interests are best defined by the patient's point of view. Proxy deciders are, then, obligated to replicate that point of view insofar as possible. But most decisions we make affect the lives of others—that is the main reason there is a body of ethics in the first place. And the present incompetence of a patient should not obligate others to perpetuate the patient's former selfish ways.

It would, of course, be possible to modify and defend the doctrine of substituted judgment by reinterpreting the concept of autonomy.[5] Patient autonomy is, after all, the main reason we embrace substituted judgment, and we usually define patient autonomy as "what the patient would have wanted." But if we were to work instead with a truly Kantian notion of autonomy, we would arrive at a very different theory of substituted judgment. For Kant would insist that a domineering, selfish person would acknowledge that she deserves less generous care when she becomes incompetent than a more caring, giving person deserves. While she might not actually elect less generous care if she were able to choose for herself, the moral judge within her would recognize that she deserves less care from others due to the way she has treated them.

On Kant's view, then, the treatment she would choose for herself is not the appropriate standard of autonomy. Rather, her judgment about what is fair or what she now deserves would be the true meaning of autonomy. Kant would insist that the selfish,

5. I owe this point to an anonymous referee.

domineering ways of an individual are all heteronomous, despite the fact that the person consistently chose them. He would further insist that a request for medical care that requires inordinate sacrifices from one's family is also heteronomous, even if the patient would have wanted that. This interpretation of autonomy and substituted judgment are clearly very different from the standard interpretation in medical ethics.

Barring a radical rethinking of the very concepts of autonomy and substituted judgment, the doctrine of substituted judgment must be rejected. At the very least, our standard view of substituted judgment must be replaced with a theory in which the interests of the incompetent are constrained by what is morally appropriate, *whether or not* the patient would have so constrained herself. Often, the patient would have been sensitive to the interests of the rest of the family, but not always. In any case, the interests of other members of the family are not relevant to proxy decisions *because* the patient would have considered them as part of her own interests, they are relevant *whether or not* the patient would have considered them.[6] It is simply not the regard of the patient for the interests of her family that give those interests moral standing. No patient, competent or incompetent, deserves more than a fair, equitable consideration of the interests of all concerned. Fairness to all includes, I would add, fairness to the patient herself in light of the life she has lived and especially the way she has treated the members of her family.

The theory of proxy decision making must be rebuilt. While proxy deciders must guard against *undue* consideration of their own interests, undue consideration of the *patient's* interests is likewise to be avoided. Proxy deciders have been given the wrong instructions. Instead of telling them that they must attempt to put themselves into the shoes of the incompetent patient and decide as she would have decided, we must tell them that the incompe-

---

6. Thus, I am in substantial disagreement with even that one paragraph from the PRESIDENT'S COMM'N REPORT, *supra* note 4, at 135-36, which goes farthest toward something like the position I embrace. The *President's Commission Report* would allow proxies to consider the interests of family members only if there is substantial evidence that the patient would have considered the family members' interests. *Id.* But in my view, this is not the reason that the interests of the members of the patient's family are relevant. If the patient was a selfish, inconsiderate person, this does not mean that the interests of her family somehow become morally illegitimate or irrelevant.

tent patient's wishes are the best way to define *her* interests, but what she would have wanted for herself must be balanced against considerations of fairness to all members of the family.

V.

Fundamental changes in the theory of proxy decisions will need to be created and defended, and a view such as mine faces a host of important questions. I cannot develop an alternative theory in this paper; indeed, I cannot even fully answer the most pressing questions about an alternative. Here, I can only provide suggestions about the way I would try to approach four of the most immediate questions about the theory of proxy decisions I would advocate.

(1) Proxy deciders must, as I have said, avoid *undue* consideration of either their own interests or the interests of the patient. But how is "undue consideration" to be defined? A full answer to this question would require an account of the family and of the ethics of the family. We can begin, however, by noting that, *prima facie*, equal interests deserve equal consideration. But what defines "equal interest?" Norman Daniels has developed the concept of a "normal opportunity range" for the purpose of allocating resources to different individuals and different age groups.[7] Perhaps this concept could be extended to problems of fairness *within* families by asking how different treatment options will affect the "opportunity range" of the various members of the family. If so, "undue consideration" could be partially defined as a bias in favor of an interest that affects someone's opportunity range in a smaller way over an interest that affects another's opportunity range in a greater way.

But even if this suggestion about the "opportunity range" could be worked out, it would represent only one dimension of an adequate account of "undue consideration." Another dimension would be fairness to competent and formerly competent members of the family in light of the way they have lived and treated each other.

(2) *Whose* interests are to be considered? For example, what about the interests of family members who do not care for the

7. DANIELS, *supra* note 3, at 36-42.

patient or who have long been hostile to the patient? Lack of concern for the patient and even hostility toward the patient do not, on my view, exclude family members from consideration. Such family members still may have important interests at stake. Moreover, we must not assume that the neglect or hostility is not merited. Family members' neglect or hostility toward the patient might, however, diminish the weight we should give their interests.

What of the interests of close friends or companions who are not members of the family? "Family," as I intend this concept, is not restricted to blood or marital relationships. Close friends, companions, unmarried lovers—all of these relationships may entitle persons to consideration in treatment decisions. Those who are distant—neither emotionally involved with the patient nor related by blood or marriage—will almost never have strong enough interests in the treatment of a patient to warrant consideration. (Health care professionals may have strong interests, but they have special professional obligations to ignore their own interests and are usually well compensated for doing so.) I see no principled way to exclude consideration of anyone whose interests will be importantly affected by a treatment decision.

(3) Would not any theory like the one I propose result in unfair treatment of incompetent patients? After all, we do not require that competent patients consider the interests of their families when making treatment decisions. And if competent patients can ignore their families, does not fairness require that we permit incompetent patients to do so as well? I have argued elsewhere that if we want to insist on patient autonomy, we must insist that patients have *obligations*, as well as rights.[8] In many cases, it is irresponsible and wrong for competent patients to make self-centered or exclusively self-regarding treatment decisions. It is often wrong for a competent patient to consider only which treatment she wants for herself. We must, then, start trying to figure out what to do when patients abuse their autonomy—when they disregard the impact of their treatment decisions on the lives of others. Sometimes, no doubt, we should seek to find ways to

8. John Hardwig, *What About the Family?*, 20 HASTINGS CENTER REP., Mar./Apr. 1990, at 5, 8-9.

prevent patients from abusing their autonomy at too great a cost to their families.

Still, competent patients are almost always permitted to ignore the interests of their family members, even if it is wrong to ignore them. We do not force them to consider the impacts of their decisions on others, nor do we disallow their decisions if they fail to do so. How, then, can it be fair to incompetent patients to develop a theory of proxy decisions that will, in effect, hold them to a more stringent moral standard by requiring them to accept treatment decisions made in light of their families' interests? The answer to this question is that there are many things that we are at liberty to do, but only so long as we do not need an agent to help us accomplish them. If we can file our own taxes, we may be able to cheat in ways that a responsible tax advisor would refuse to do. We may get away with shoddy deals that an ethical lawyer would not be a party to. Thus, the greater freedom of competent patients is only a special case of the generally greater freedom of action when no assistance of an agent is required.

(4) What about the legal difficulties of an alternative view of proxy decision making? They are considerable; it is presently *illegal* to make proxy decisions in the way I think is morally appropriate. The courts that have become involved in proxy decisions have almost all opted for exclusively patient-centered standards. I do not have the expertise needed to address the legal issues my view raises. My purpose here is to challenge the faulty moral foundations which undergird present legal practice.

It is possible that family law could provide a model for a revised legal standard of proxy decision making. Family law recognizes the legitimacy of proxy decisions—for children, for example—that are not always in the best interest of the person represented by the proxy. It must, if only because there are many cases in which the interests of one child will conflict with those of others. Nor does family law require parents to ignore their own interests in deciding for a child; instead, it defines standards of minimum acceptable care, with the hope that most families will do better than these minimum standards. Perhaps we should similarly separate the legal from the moral standard for proxy decisions. If no abuse or neglect is involved, the legal standard is met, though that may be less than morality requires of a proxy decision maker.

VI.

All these issues—about undue consideration, about eligible interests, about fairness between competent and incompetent patients, about the law of proxy decisions—may seem very complex. I do not believe they are unnecessarily complicated. Many important decisions within families are very complicated. In medical ethics, we have simplified our task by working with an artificially over-simplified vision of the interests and decisions of families in medical treatment. So, if my critique of the present theory of proxy decisions is correct, we all—medical ethicists, reflective health care practitioners, legal theorists, and lawyers—have a lot of hard work to do. The change I propose is basic, so the revisions required will be substantial.

I close now with a word of caution and a word of encouragement. The word of caution: We must recognize that even the necessary revisions in our moral and legal theories of proxy decisions would not resolve all the problems of proxy decisions. Proxy deciders with interests that conflict with those of the patient do face serious moral difficulties and very real temptations to give undue weight to their own interests. Although the concepts of both "overtreatment" and "undertreatment" will have to be redefined in light of the considerations I have been advancing, pressures from proxies for inappropriate treatment will remain. I do not wish to minimize these difficulties in any way.

But we should not give proxies the morally erroneous belief that their own interests are irrelevant, then censure them for allowing their interests to "creep in" to their decisions. Instead, we must deal forthrightly with the very real difficulties arising from interested proxy decisions, by making these interests conscious, explicit, and legitimate. Then we must provide guidance and support for those caught in the moral crucible of proxy decisions. Not only would this approach be ethically sounder, it would decrease the number of inappropriate proxy decisions.

Finally, an encouraging word. The Alzheimer's case that I have cited notwithstanding, the practice of medicine is often better than our ethical theories have been. It has generally not been so insensitive to the interests of family members as our theories would ask that it be. Indeed, much of what now goes on in intensive care nurseries, pediatrician's offices, intensive care units,

and long-term care facilities makes ethical sense *only* on the assumption that fairness to the interests of the other members of the family is morally required. To mention only the most obvious case, I have never seen a discussion about institutional versus home care for an incompetent patient that did not attempt to address the interests of those who would have to care for the patient, as well as the interests of the patient.

Nevertheless, current ethical theory and traditional codes of medical ethics can neither help nor support health care professionals and proxies struggling to balance the patient's interests with those of the proxy and other family members. Indeed, our present ethical theory can only condemn as unethical any weight given the interests of the family. Thus, our ethical theory forces us to misdescribe decisions about institutionalization in terms of what is physically or psychologically *possible* for the family, rather than in terms of what is or is not too much to *ask* of them. If we were to acknowledge the moral relevance and legitimacy of the family's interests, we would be able to understand why many treatment decisions now being made make sense and are not unethical. We would then be in a position to develop an ethical theory that would guide health care providers and proxies in the throes of excruciating moral decisions.

Bioethics and the Family: The Cautionary View from Family Law*

*Carl E. Schneider***

I. INTRODUCTION

For many years, the field of bioethics has been specially concerned with how the authority to make medical decisions should be allocated between doctor and patient. Today the patient's power—indeed, the patient's right—is widely acknowledged, at least in principle. But this development can hardly be the last word in our thinking about how medical decisions should be made. For one thing, sometimes patients cannot speak for themselves. For another, patients make medical decisions in contexts that significantly include more participants than just the patient and doctor. Now, as this conference demonstrates, bioethics is beginning to ask what role the patient's family should play in making medical decisions.

In addition, bioethics has in recent years increasingly been required to address another kind of problem: How should we resolve the ethical dilemmas associated with matters of reproduction—particularly novel means of reproduction, like *in vitro* fertilization and surrogate motherhood? As the technical capacities of medicine have expanded, these bioethical questions have raised

---

** Professor of Law, University of Michigan. This Article is a somewhat altered version of a paper presented at the University of Utah College of Law and *Utah Law Review* Symposium on Ethics, Bioethics, and Family Law. An earlier incarnation of that paper was presented at a conference at the Hastings Center on the Family and Bioethics. I am grateful to the conferees at both the University and the Hastings Center for their helpful comments and to both institutions for providing an atmosphere so admirably conducive to the rational discussion of controversial questions. Finally, I am glad to thank my colleague Patricia D. White for her insightful responses to an earlier draft of this manuscript.

For the reasons described in Richard A. Posner, *Goodbye to the Bluebook*, 53 U Chi L Rev 1343 (1986), I will adhere to *The University of Chicago Manual of Legal Citation* (Lawyers Co-Op, 1989). As a great man once said, "*Faites simple*." I am enthusiastically grateful to the editors of the *Utah Law Review* for the uncommonly generous way in which they have accommodated my wish to strike a blow for freedom from the formalisms and fatuities—no, the inanities and insanities—of *The Bluebook*.

pressing and puzzling issues about what a family is and how it should and should not be created.

In short, bioethics today confronts ultimate and essential questions about the ethical and social bases of family life. My task here is to ask what bioethics might learn about these troubling questions from the experience of another field which has wrestled with them for centuries—family law. I have a second task as well. This is, after all, a symposium on family law, and I hope that I might make myself useful to that field by providing in brief form and with concrete illustrations a taxonomy and survey of some of family law's basic conceptual approaches.

Family law *ought* to have something to say to bioethics about these problems. To begin with, many bioethical issues directly concern family law. For instance, family law seeks to regulate the situation in which children are created and given families. Many other bioethical issues deal with matters—like decisions about medical care—that impinge on family life and that family law has thus been interested in. More generally, family law has long experience with a multitude of ethical problems involving the relations between and the regulation of family members. Family law therefore should have developed vocabularies and approaches that could illuminate bioethical problems.

At the very least, we might expect family law to offer bioethics some concept of the family. Family law ought to have developed some definition of "family," since it needs to know what it is about a grouping of people that makes a grouping a family. It ought also to have reached some understanding about what the moral and social relationships of family members are, so that it can know what claims they may make on each other and what duties they owe each other. Such a conception of the family is surely crucial to both major branches of bioethics, since they deal precisely with the creation of families and often with the responsibilities of family members.

My paper will be divided into several sections, each devoted to a particular conceptual approach to the problems of family law. Each section will describe the approach, briefly evaluate its current status in family law, and then ask what usefulness the approach might have for bioethical problems. Ultimately I will

argue that family law offers no vocabulary or approach that can directly and readily be adopted in analyzing bioethical issues. This conclusion should not be surprising. Family law is law operating at its outer limits, trying to govern the most ungovernable of human relationships, seeking to understand the most mysterious and controversial aspects of social life.[1]

Yet all this does not mean that bioethics can learn nothing by looking at family law. On the contrary, there is much to be gained by looking at the reasons for family law's conceptual limits and practical constraints; for those reasons reveal something about the claims, conflicts, and contradictions that make bioethical questions so painful and contemporary family law so problematic.

II. MORAL DISCOURSE

By definition, bioethical problems raise moral issues. What kind of discourse about moral issues does family law use and how might that discourse be recruited to deal with bioethical issues? Until recently, we might plausibly have tried to answer that question, since over the preceding century family law had developed a tolerably clear definition of the family and a reasonably coherent body of beliefs about the relations among family members and the purposes of family life. That definition and those beliefs had a fairly well understood moral basis, they were articulated in moral terms, and they required courts to analyze many individual cases at least partly in moral language. In the last two or three decades, however, family law has increasingly eschewed moral discourse. That is, there has been (with some exceptions) growing reluctance to have the law serve expressly moral goals, to articulate legal principles in moral terms, and to have courts analyze problems in moral language. Simultaneously, many moral decisions have been transferred from the law to the

1. For an investigation of these features of family law, see Carl E. Schneider, *The Next Step: Definition, Generalization, and Theory in American Family Law*, 18 U Mich J L Ref 1039 (1985).

people the law once sought to regulate.[2] No-fault divorce exemplifies this trend:

> [B]efore no-fault divorce, a court discussed a petition for divorce in moral terms; after no-fault divorce, such a petition did not have to be discussed in moral terms. Before no-fault divorce, the law stated a view of the moral prerequisites to divorce; after no-fault divorce, the law is best seen as stating no view on the subject. Before no-fault divorce, the law retained for itself much of the responsibility for the moral choice whether to divorce; after no-fault, most of that responsibility was transferred to the husband and wife.[3]

The waning of moral discourse in family law has a number of causes, including the doctrine of family autonomy,[4] the tradition of liberal individualism, a series of modern upheavals in moral beliefs, the constitutionalization of family law, and the medicalization—especially the "psychologization"—of social issues. But several of the trend's most central causes can be summarized by the phrase "the standards problem." An important justification for the doctrine of family autonomy has long been that people disagree about how families ought to be organized and run and that those disagreements often reduce to unresolvable disputes over unverifiable beliefs. Americans have grown increasingly sensitive to cultural and individual variations in views on these subjects and have increasingly felt that society should not impose its standards on people, particularly where those standards affect people's intimate relations. For all these reasons, it is increasingly felt that standards for governing family relations cannot and should not be written.

---

2. For a full statement of this hypothesis, see Carl E. Schneider, *Moral Discourse and the Transformation of American Family Law*, 83 Mich L Rev 1803 (1985). In that Article, as here, I argue only that the developments I describe are a trend, not a fully accomplished fact. For a (partial) normative evaluation of the trend, see Carl E. Schneider, *Rethinking Alimony: Marital Decisions and Moral Discourse*, 1991 BYU L Rev 197, 233-57. For characteristically thoughtful comments on the trend, see Lee E. Teitelbaum, *Moral Discourse and Family Law*, 84 Mich L Rev 430 (1985).

3. Schneider, 83 Mich L Rev at 1810 (cited in note 2).

4. This is the standard principle of family law that the state ought wherever possible to refrain from "intervening" in the family. For a discussion of what "intervention" might mean, see Schneider, 1991 BYU L REV at 235-43 (cited in note 2).

What consequences do the diminution in moral discourse and the standards problem which partly underlies it have for family law's usefulness to bioethics? One consequence has been that a plausible (if not always optimal) means of resolving such questions—directly addressing the moral issue presented by the bioethical issue—is made less attractive or is even foreclosed. *Roe v. Wade* exemplifies this point in two ways. First, *Roe*'s holding largely removed from the law's purview the issue of the morality of abortion in general and of any individual's abortion in particular. Second, the opinion's reasoning expressly sought to reach a conclusion without discussing the morality of abortion. Less dramatically, the diminution in moral discourse and the standards problem have meant that family law's cupboard is increasingly bare of moral concepts of the family that might inform discussions of such bioethical dilemmas as surrogate-mother agreements and of the role families should play in making medical decisions for their incompetent members.[5] There is in fact some evidence that courts directly confronting bioethical problems have sought to do so without embarking on moral inquiries. As Allen Buchanan notes, for instance, "From *Quinlan* on, the courts have attempted to avoid the fundamental philosophical and constitutional issues raised by the task of developing a more adequate concept of the person and hence of the death of a person."[6]

For us, however, the most momentous consequence of the trend away from moral discourse in the law and of the unremitting prominence of the standards problem has been that the law is more and more driven to find ways around the standards problem. That is, the law has increasingly had to ask, if we cannot directly address the moral aspects of the issues we face, what other ways can we find of analyzing and resolving them? The rest of this paper will examine some of the leading alternatives.

---

5. It is worth observing that there are institutional differences in the willingness to engage in moral discourse. Such discourse is likeliest to occur in legislatures, partly because the need to write statutes and pressure from constituents and interest groups often bring moral issues to the fore.

6. Allen E. Buchanan, *The Limits of Proxy Decision-Making*, in Rolf Sartorius, ed, *Paternalism* 153 (U Minn Press, 1983).

III. THE PROTECTIVE FUNCTION

One way out of the dilemmas caused by the diminution of moral discourse and by the standards problem has been to justify governmental action in terms of one of family law's least controversial enterprises—the protective function. That function effectuates the law's duty to protect citizens against the various harms that might befall them, and particularly to protect them from injuries done them by other citizens.[7] Moral discourse in family law is presently strongest and the standards problem is presently weakest in those areas where it can be said that the law is protecting someone who cannot protect himself, who is helpless against a more powerful person. Thus some of the topics in family law most often discussed today in moral terms are spouse abuse, child abuse, and child support. In other words, the protective function can sometimes seem uncontroversial enough or pressing enough to escape some of the strictures of the standards problem.

Family law's protective function might seem to offer useful approaches at least to some bioethical problems. Thus it is sometimes said that surrogate-mother contracts ought to be prohibited in the interest of protecting surrogates from the pains of having to give up a child who is (often) genetically theirs and (always) gestationally theirs. And thus it is sometimes suggested that the ability of parents to refuse medical treatment for their defective newborn infants ought to be supervised and superseded in order to protect those infants.

But the protective function is subject to (at least) four generic problems which, in bioethics as in family law, will often prove disabling. The first is that protection easily degenerates into paternalism: It will often seem improper to protect people who do not want protection or who even actively resist it. It was this fear, for example, that in important part motivated Justice Brennan's dissent in *Cruzan v. Director, Missouri Department of Health*.[8] He

---

7. For an extended treatment of the protective function, see Carl E. Schneider, *Family Law: Cases and Materials* (West, forthcoming 1993).

8. 58 USLW 4916 (1990). Nancy Cruzan was a young woman who had fallen into a persistent vegetative state after an automobile accident. Her parents sought to have the hospital in which she lay discontinue her food and water. A Missouri statute, however, required anyone asking that food and water be withheld from a patient in a persistent vegetative state show by clear and convincing evidence that that withdrawal was what

argued that the state was protecting Nancy Cruzan's life, but that she found that life a burden, wished to end it, and was entitled to do so.[9] Similarly, a blanket prohibition of surrogate-mother contracts could be justified as protecting women from the misery of losing children they had borne and wished to keep. But such a prohibition would be regarded by many prospective surrogates as an undue interference with their liberty and an inaccurate reflection on their ability to make decisions for themselves.

A second generic problem of the protective function is that it will not always be clear what "protection" in a given case means. Was Nancy Cruzan being protected by the state, which wished to preserve her life, or by her parents, who wished to save her from what her life had become? Would statutes prohibiting abortion protect the lives of unborn children? Or does *Roe v. Wade* protect pregnant women from the dangers of abortion statutes? As these questions are intended to suggest, attempts to serve the protective function can return us to the standards problem and to its underlying issues about what makes life good, matters as to which answers are obscure and agreement is elusive.

The protective function's third characteristic problem is that, in trying to protect people from one harm, the state—because it is large, complex, cumbersome, and obliged to follow rules that must often be broadly phrased and inflexibly interpreted—will sometimes, perhaps frequently, injure people in unanticipated ways. Worse, the injured people can easily be those the law is most anxious to help. For instance, we might want to judicialize medical decisions in order to protect patients from improvident decisions to terminate treatment. But the classic defect of such judicialization is that it imposes painful burdens in time, trouble, expense, and misery on doctors, nurses, families, and, what is worst, on the patients themselves.

The fourth generic problem with the protective function is that sometimes the law cannot safeguard all the people who may seem to need help because their interests conflict. In a surrogate-mother case, do you protect the surrogate, whose deep attachment

---

the patient would have wanted. The Missouri courts held that Cruzan's parents had not made such a showing. The United States Supreme Court held that nothing in the United States Constitution prevented Missouri from imposing such a requirement. I discuss *Cruzan* and the rights thinking that undergirds it in some detail in Part VI.

9. Id at 4926-34 (Brennan dissenting).

to her child is threatened? Do you protect the contracting parent, for whom surrogacy may be the only way of having a biologically related child and who has nurtured months of expectations and hopes? Do you protect the infant, the one person in the story who cannot speak for himself?

In sum, the flaw of the protective function as a path of escape from the standards problems is that it works best in the easy cases. In poorly explored and daunting areas like the bioethical conflicts we are discussing, resorting to the protective function as justification is likely only to force us back to those moral questions we had hoped to escape.

IV. OFFICIAL DISCRETION

When the law finds itself unable to write standards, it often transfers decisions to the discretion of an official or a judge. This is an old technique in the law generally, and for excellent reasons. Courts and bureaucracies often need flexibility to adjust their decisions to the world's complexity. Judges and administrators are frequently accorded discretion because would-be rule makers realize that they cannot anticipate all the circumstances in which they might wish a rule to be applied, because they hope that judges will be well-situated to construct rules by accretion as they gain experience deciding cases in an area, and simply because rule makers find they cannot agree on a rule.[10]

According judges discretion is, of course, a recurring family-law technique for avoiding direct confrontations with the standards problem. A particularly vivid example of the technique in that field is the law of child custody, which uses the markedly discretionary criterion of the child's "best interest." It is a technique which has found fresh favor in the law governing the allocation of a couple's property on divorce, in which courts may now be directed to divide the spouses' property "equitably."

Despite the regularity with which family law has substituted discretion for standards, the technique is not in good odor in the field. Virtually every major figure in the field has condemned

10. For a more extensive survey of the merits and demerits of rules and discretion, see Carl E. Schneider, *Rules and Discretion: A Lawyer's View*, in Keith Hawkins, ed, *The Uses of Discretion* (Oxford U Press, forthcoming 1993).

child-custody law's best-interest standard as deplorably indeterminate.[11] Equitable distribution *simpliciter* is not well-established and is currently under attack. Federal law now calls for mechanical guidelines to replace discretionary awards of child support. And there is notable sentiment in favor of substituting elaborately specific criteria for intervention in families in child-abuse-and-neglect cases for the old and discretionary intervene-whenever-it's-necessary standard.

The reasons official discretion is unloved in family law largely apply to bioethics. These reasons are too familiar to bear prolonged reiteration here, since they are the standard objections to discretion. They include the arguments that discretion allows officials and judges to let their prejudices affect their decisions, that discretion leads to inconsistent decisions, and that discretionary standards give affected parties insufficient guidance about what the law expects of them or will do to them. Further, granting officials and judges discretion solves the standards problem only in the sense of relieving a legislature of the tasks of formulating, articulating, and getting the votes to enact standards. After all, an official or judge must base a decision on some kind of principle, even if it is unarticulated or even unconscious. Discretion does not eliminate the question whether the principle chosen is a good one and whether it is right to hold people to it rather than allowing them to choose for themselves how to behave. These kinds of problems with discretion may be made more concrete by imagining what the Court's reaction to confiding a decision in *Cruzan* to the unfettered discretion of a judge or official would have been. Many of the Justices, at least, would have protested that such a rule grievously violates a patient's rights to decide what care to receive and to enjoy the benefits of due process.[12] Thus, while awarding

---

11. For two first-rate examples of those criticisms, see Robert H. Mnookin, *Child-Custody Adjudication: Judicial Functions in the Face of Indeterminacy*, L & Contemp Probs 226 (Summer 1975); David L. Chambers, *Rethinking the Substantive Rules for Custody Disputes in Divorce*, 83 Mich L Rev 477 (1984). For a critical review of those criticisms and a cautious and constrained defense of discretion, see Carl E. Schneider, *Discretion, Rules, and Law: Child Custody and the UMDA's Best-Interest Standard*, 89 Mich L Rev 2215 (1991).

12. There are, however, many forms of official discretion, and it is often extremely difficult to formulate rules for resolving complex problems without confiding a good deal of discretion in some official. The opinions in *Cruzan* generally seem to contemplate that, at least in many instances, a court would have to decide whether an incompetent patient would have wanted treatment terminated. Given the probable quality of evidence in

grants of discretion may in fact be a good way of handling many bioethical problems, family law is at best uneasy authority for that conclusion.

V. FAMILY AUTHORITY

If the state cannot promulgate standards, and if it cannot finesse the problem away by giving officials or judges discretion, it must transfer decisions to someone else. As I said earlier in my discussion of moral discourse, this is exactly what family law has tried to do. Even before the "transformation" of family law, numerous decisions were assigned to "the family." Thus courts have long refused to resolve many kinds of disputes between family members on the grounds that families ought to be encouraged to work out their own problems in their own way. And thus states have long confided responsibility for most decisions about children to their parents.

The application of this view to a number of bioethical issues is obvious and appealing. Some of the bioethical decisions associated with reproduction already have been or might plausibly be resolved by referring them to the family. Thus decisions about the morality of an abortion have been transferred to the pregnant woman, in part with the expectation (but not compulsion) that she will share that decision with the father. Decisions about medical care for incompetents are, at least in practice, often made by the patient's family,[13] and many people believe, as the dissents in *Cruzan* indicate, that this is right and proper.[14] Nevertheless, family law's experience suggests some difficulties with solving the standards problem by deferring to "the family."

---

many of these cases, this is a decision which it will often be impossible to make without a considerable exercise of discretion.

13. See, for example, Stewart B. Levine, et al, *Informed Consent in the Electroconvulsive Treatment of Geriatric Patients*, 19 Bul Am Acad Psych L 395 (1991); Clara C. Pratt, et al, *Autonomy and Decision Making Between Single Older Women and Their Caregiving Daughters*, 29 Gerontologist 792 (1989). In practical fact, of course, the family's power is, at best, shared with physicians. For an illuminating investigation of the relationship between doctors and families in medical decisions, see Robert Zussman, *Intensive Care: Medical Ethics and the Medical Profession* (U Chi Press, 1992).

14. For an influential statement of this position, see Nancy K. Rhoden, *Litigating Life and Death*, 102 Harv L Rev 375 (1988).

Perhaps the most basic of these difficulties is that the conceptual basis for this deference has been eroded in recent years. Traditionally, as I suggested a moment ago, the law assumed that family members are united by bonds of mutual concern so strong that the law could and should treat each family as a whole and not just as a collection of individual family members.[15] Thus the law was willing to have family members make decisions for each other and even (as I suggested above) to insist that decisions be made within families rather than by courts.

Increasingly, however, courts and commentators have attacked this view of the family. To some critics, deferring to "the family" really means confirming the power of its most powerful member.[16] To some of these critics, such deference simply affirms the patriarchal principle. To others, it denies the rights and personhood of children. To still others, it threatens the autonomy and self-sufficiency of all members of the family. Yet other critics find the entity view simply mistaken, on the reasoning that families are irreducibly made up of individuals and have no interests other than those of their members. Finally, critics who are concerned about the standards problem find the entity view objectionable because it embodies and promotes a normative ideal—however vague—of the family.

Family law has thus more and more come to regard family members as individuals who no doubt have important relationships with each other but who should be treated as legally distinct. As Justice Brennan wrote in a telling and influential passage, "[T]he marital couple is not an independent entity with a mind and heart of its own, but an association of two individuals each with a separate intellectual and emotional makeup."[17] Professor Hafen describes the new view of the family as contractual. He suggests that families are increasingly united by a merely contractual solidarity, one whose "main motivation is 'purposive,

---

15. For a thoughtful statement of this view, see Bruce C. Hafen, *The Family as an Entity*, 22 UC Davis L Rev 865 (1989).

16. See, for example, Lee E. Teitelbaum, *Family History and Family Law*, 1985 Wis L Rev 1135; Frances E. Olsen, *The Myth of State Intervention in the Family*, 18 U Mich J L Ref 835 (1985). For comments on this view, see Schneider, BYU L Rev at 235-43 (cited in note 2).

17. Eisenstadt v. Baird, 405 US 438, 453 (1972).

implicitly egoistic, utilitarian,' and lacking in a 'sense of sociocultural oneness of the parties.' Each party typically enters the relationship 'for his own sake, uniting with the other party only so far as this provides him with an advantage (profit, pleasure, or service).'"[18]

The law now recognizes the individuality of family members in a variety of ways. For example, no-fault divorce, by making divorce available on demand, forswears any legal effort to hamper each spouse's ability to leave the family. Family law has increasingly allowed spouses (and unmarried cohabitants) to contract with each other. The ever-more-common practice of appointing lawyers to represent children in cases—like custody or medical-care proceedings—in which their parents are litigants further recognizes the legal separateness of family members and the possible (or even presumptive?) adversity of their interests.[19] This pattern similarly presents itself in cases raising bioethical issues. *Bellotti v. Baird*[20] makes one important kind of decision—whether a child should have an abortion—essentially a decision for the child alone, and not one for the family.[21] *Planned Parenthood of Central Missouri v. Danforth*[22] makes the wife's decision whether to have an abortion one she may make without obtaining her husband's consent.[23]

We have looked at the way family law has tried to escape the standards problem by referring decisions to the family. But, I have been arguing, the standards problem itself (at least in some of its more sweeping versions) undercuts the basis for any such referral.

---

18. Hafen, 22 UC Davis L Rev at 895-96 (cited in note 15.)

19. See, for example, *Michael H. v. Gerald D.*, 491 US 110 (1989), where a wife allegedly bore the child of a man other than her husband. Id at 113-14. The child's putative natural father sued to be declared the legal father. Id at 118. When the child was a little less than two years old, the court appointed a lawyer to represent her. Id at 114. When the Supreme Court decided the case the child was slightly over eight years old, and she was still represented by counsel. Id at 130-32. For an examination of the case in light of the state's interest in the family as an entity, see Carl E. Schneider, *The Channelling Function in Family Law*, 20 Hofstra L Rev 495 (forthcoming 1992). For investigations of the problems posed when lawyers represent clients who cannot speak for themselves, see Robert H. Mnookin, et al, *In the Interest of Children: Advocacy, Law Reform, and Public Policy* (WH Freeman & Co, 1985); and Carl E. Schneider, *Lawyers and Children: Wisdom and Legitimacy in Family Policy*, 84 Mich L Rev 919 (1986).

20. 443 US 622 (1979).

21. Id at 643-44.

22. 428 US 52 (1976).

23. Id at 69.

The referral is best justified by a moral view of the family that makes it an appropriate decision maker. The law once accepted such a view. Now, partly because of the standards problem, it is disinclined to do so. Thus the rationale for deferring to the family is markedly weakened.[24]

The problem with referring bioethical decisions to the family is not just that the basis for such a referral has been eroded. It is also that, unless "family" is quite broadly defined, some number of people will have no family to which a decision can be referred (and even if it is broadly defined, some people will still not have a family able and willing to take on the burden of their bioethical decisions). For example, within the ordinary understanding of the law, a single adult like Nancy Cruzan[25] has been emancipated from her family, so that her parents can no longer make decisions for her. But why not simply define Nancy Cruzan's family to include her parents? In her case, and in many cases, that is no doubt the right thing to do. But whatever the wisdom of defining "family" broadly in particular cases, family law may hesitate to do so when writing generally applicable rules. Let us ask why.

The family to which decisions are ordinarily referred is essentially the nuclear family.[26] Within it are two kinds of relationships. There are special reasons to expect people in each of these relationships to make good decisions for each other. The first relationship is marital. Husbands and wives should make good decisions for each other because they have confided special love and trust in each other. The second relationship is parental. Parents should make good decisions for their children because of the perhaps-instinctual feelings of love, concern, and responsibility parents have for their young children. The quality of both kinds of decisions should be enhanced where families live in households: People who are committed to living intimately together for many

---

24. I use "weakened" advisedly. The entity view of the family retains many adherents even in the remotenesses of academe. See, for example, Hafen, 22 UC Davis L Rev (cited in note 15). It is probably the conventional wisdom among much of the rest of the country. See Carl E. Schneider, *State-Interest Analysis in Fourteenth Amendment "Privacy" Law: An Essay on the Constitutionalization of Social Issues*, L & Contemp Probs 79, 107-110 (Winter 1988). And even people who reject it will still often find reasons in particular cases to prefer familial to governmental decisions.

25. It appears that her marriage ended sometime after her accident. *Cruzan*, 58 USLW at 4917.

26. See, for example, *Michael H.*, 491 US at 117-30.

years should act wisely for each other because they come to know each other deeply and because their interests become so richly intertwined.[27]

Of course, parents are still, today as yesterday, bound by ties of blood, love, and experience to their adult children. But more and more, American society expects parents to raise their children to develop their own values, to leave their parents' homes, to establish their own households, to lead their own lives. "Therefore shall a man leave his father and his mother, and shall cleave unto his wife: and they shall be one flesh."[28] Partly in pursuit of that goal, family law has increasingly promoted the autonomy of minor children. *A fortiori*, will it not promote the autonomy of adult children?[29] If the idea that parents of an adult child should not be legally considered part of the child's family seems plainly wrong, consider the cases we now sometimes see in which parents of an incompetent adult contend with his homosexual lover for the power to make decisions for him.

I have been discussing the problems with broadening the definition of family in its easiest form—to include parents of adult

---

27. This argument, however, should not be pushed too far. There is evidence that many people do not discuss their preferences about medical care with their families. See, for example, Dallas M. High, *All in the Family: Extended Autonomy and Expectations in Surrogate Health Care Decision-Making*, 28 Gerontologist 46 (Supp 1988); Bernard Lo, et al, *Patient Attitudes to Discussing Life-Sustaining Treatment*, 146 Archives Internal Med 1613 (1986). Worse, there is some direct reason to doubt that families in fact accurately learn their members' preferences about medical care from living with them. See, for example, Tom Tomlinson, et al, *An Empirical Study of Proxy Consent for Elderly Persons*, 30 Gerontologist 54 (1990); Joseph G. Ouslander, et al, *Health Care Decisions Among Elderly Long-Term Care Residents and Their Potential Proxies*, 149 Archives Internal Med 1367 (1989); Allison B. Seckler, et al, *Substituted Judgment: How Accurate Are Proxy Predictions?*, 115 Annals Internal Med 92 (1991); Richard F. Uhlmann, et al, *Physicians' and Spouses' Predictions of Elderly Patients' Resuscitation Preferences*, 43 J Gerontology M115 (1988).

Despite all this, there is also evidence that people generally prefer that their families participate in making medical decisions while they are competent and make medical decisions for them when they cannot do so themselves. See, for example, Madelyn A. Iris, *Guardianship and the Elderly: A Multi-Perspective View of the Decisionmaking Process*, 28 Gerontologist 39 (Supp 1988); Dallas M. High and Howard B. Turner, *Surrogate Decision-Making: The Elderly's Familial Expectations*, 8 Theoretical Med 303 (1987).

28. Genesis 2:24. See Leslie Francis, *The Roles of the Family in Making Health Care Decisions for Incompetent Patients*, 1992 Utah L Rev 861.

29. For insightful reflections on the autonomy and responsibility of adult children, see Lee E. Teitelbaum, *Intergenerational Responsibility and Family Obligation: On Sharing*, 1992 Utah L Rev 765; Hilde L. Nelson and James L. Nelson, *Frail Parents, Robust Duties*, 1992 Utah L Rev 747.

children. All these problems worsen when we consider broadening the definition to include people unrelated by blood. What may we infer from family law's experience with this problem? The law has become more willing to treat as a "family" a group of people whose relationship performs the functions that a traditional family performs. Thus *Marvin v. Marvin*[30] offers protections to people leaving non-marital cohabitation that resemble those offered to people ending marriages.[31] And thus *Braschi v. Stahl Associates*[32] treats a homosexual couple as a family for purposes of New York City's rent control program.[33]

But this "functional equivalence" approach has its difficulties and drawbacks. Thus courts have hesitated to extend *Marvin*'s principle to reach other ways of treating cohabitants like spouses. They have done so for reasons that are relevant to our inquiry. Marriage represents a specially serious and binding commitment two people make to each other. That commitment forms the basis for treating spouses in special ways. Of course, people don't have to marry in order to make such commitments, and some unmarried couples may be as deeply and solidly bound as any husband and wife. But unless people go through the public affirmation of the commitment that marriage constitutes, the law cannot know that they have made it. The law could, of course, inquire into the quality of each non-marital commitment to see whether it met "marital standards." And indeed *Marvin* calls for just such inquiries.[34] But they seem a distasteful invasion of privacy. Nor is it clear what standards and evidence would be used in evaluating the quality of a commitment.

In addition, there is a slippery-slope problem, a problem created by the way common-law courts tend to use precedent. Marriage, I have been arguing, provides what lawyers call a bright-line rule. It is easy to tell when a couple is married, and the law treats them as married whatever the true nature of their

---

30. 557 P2d 106 (Cal 1976).
31. Id at 116, 122-23.
32. 543 NE2d 49 (NY 1989).
33. Id at 53-55.
34. The *Marvin* court anticipated investigations into whether the parties' sexual relations were a severable part of their contract, into whether the parties had tried to avoid a marital relationship, and into any facts that might form the basis for any kind of equitable relief.

emotional relationship is. But once a court starts asking whether a non-marital relationship is the functional equivalent of marriage, it starts a process in which it compares the case at hand with the weakest case in which a couple has been found to have achieved the functional equivalent of marriage. The case at hand will sometimes seem just close enough to the weakest precedent to justify saying that it qualifies as the functional equivalent of marriage. That case then becomes a precedent, and the process begins again. The process is partly driven by the dynamics of our system of precedent. And it is accelerated by the fear that to refuse to call a relationship the equivalent of marriage is to deny its importance to the parties and is thus to demean the relationship.

Why might this process be a problem? At the end of the day lies the risk that extending the regime of functional equivalents will tend to assimilate relatively transient and shallow relationships to marriage. Yet the usefulness of marriage as a social institution depends in significant part on people's understanding that it is special, and that it is special in the seriousness of commitment that it demands. The risk, in other words, is that extending the regime of functional equivalents will erode the special qualities of marriage and reduce marriage to just one more "life style choice."[35]

This brings us to our next reservation about functional-equivalence approach. Marriage is not just an outward sign of inward commitment. It is a social and legal institution which reinforces that commitment. People who marry assume a role which carries social expectations with it, expectations most people have to some degree internalized and which are not avoided without cost. This increases the likelihood—although only the likelihood—that the quality of the relationship justifies according one member the power to make momentous decisions for the other.[36]

I have been discussing some of the knotty problems of definition that would need to be untied before deference to family authority could be fully useful in solving bioethical problems. But

35. For more complete comments on the "functional-equivalence" approach, see Schneider, 20 Hofstra L Rev (cited in note 19).

36. For an extended exposition of the role and value of social institutions in the familial realm, see id.

there is a final problem with deferring to the family, however it is defined.[37] Many of the bioethical decisions we might ask families to make are enormously consequential. They are literally questions of life and death. Yet two factors (at least) can make it hard for people in intimate relationships to decide them wisely. First, love is not the only strong feeling to which intimacy gives rise. Love can be mixed with equally potent but harsher feelings of jealousy, resentment, and even hate. Second, people in such relationships may have conflicts of interest that inhibit dispassion and diminish wisdom. Those conflicts include even the crassest kind of wish that one's relative should die so that one can receive an inheritance.[38] Less drastically and more sympathetically, they include concerns that one relative's lingering illness is damaging the well-being of

---

37. Of course, family law has not burdened itself by trying to adopt a single definition of the family. Rather, it has adopted different definitions for different purposes.

38. This is, as I say, the crassest way in which a family member's decisions might be distorted. But that does not mean that only the crassest people will be influenced by it. Archdeacon Grantly is not a bad man, but it is only after "he thought long and sadly, in deep silence, and then gazed at that still living face" that he "at last dared to ask himself whether he really longed for his father's death" so that he might be appointed to his father's bishopric. Anthony Trollope, *Barchester Towers* 12 (Doubleday, nd). In addition to the fact of this distorting motive is the fear that it arouses. Prince Hal apparently is speaking the truth when he explains that he took the crown from his father's pillow because "I never thought to hear you speak again." But the king retorts,

> Thy wish was father, Harry, to that thought:
> I stay too long by thee, I weary thee.
> Dost thou so hunger for mine empty chair
> That thou wilt needs invest thee with my honours
> Before they hour be ripe?

And he has already expostulated:

> See, sons, what things you are!
> How quickly nature falls into revolt
> When gold becomes her object!
> For this the foolish over-careful fathers
> Have broke their sleep with thoughts, their brains with care,
> Their bones with industry;
> For this they have engrossed and piled up
> The canker'd heaps of strange-achieved gold;
> For this they have been thoughtful to invest
> Their sons with arts and martial exercises:
> When, like the bee, culling from every flower
> The virtuous sweets,
> Our thighs pack'd with wax, our mouths with honey,
> We bring it to the hive; and, like the bees,
> Are murder'd for our pains.

William Shakespeare, *The Second Part of King Henry the Fourth*, act IV, sc. v.

other family members. Such concerns presumably contributed to the majority's reluctance in *Cruzan* simply to hand over treatment decisions to the family.[39] And such concerns have helped motivate family law's long-standing reluctance to cede families complete control over decisions for their members. It is worth remembering, for example, that the law of child abuse requires parents to provide needed medical care for their children, and that the criminal law is in principle prepared to punish as homicide any failure to do so that results in a child's death.[40]

Obviously, I am not arguing that families should not participate in making bioethical decisions for their members. My inclinations are quite to the contrary. But I think that the lesson of family law's experience is once again cautionary. Familial decisions can be acutely troublesome and troubling. Writing rules to govern such decisions is not without its complexities and even its dangers. Further, the atomizing tendencies of the age and its law—the ever-sharpening urge to treat family members as independent of each other—conflict harshly with the desire to confide crucial bioethical decisions to families. In short, what we have been calling "family authority" offers only partial and problematic solutions to the kinds of bioethical issues we are discussing.

VI. RIGHTS DISCOURSE

Yet another common and conventional answer to the standards problem is to analyze issues in terms of rights.[41] Rights solutions confer on rights holders the power to resolve ethical questions and thereby relieve the state of the burden of doing so. Since rights thinking is one of the dominant modes of

39. See *Cruzan*, 58 USLW at 4922. The Court also noted that "there is no automatic assurance that the view of close family members will necessarily be the same as the patient's would have been had she been confronted with the prospect of her situation while competent." Id.

40. For a challenging and illuminating illustration of the difficulty of family decisions in even the most benign circumstances, see John Hardwig, *The Problem of Proxies With Interests of Their Own: Toward a Better Theory of Proxy Decisions*, 1992 Utah L Rev 803.

41. When I say "rights discourse," I will be referring primarily to the discourse about constitutional rights in the United States today. For an analysis of that discourse and its use in family law, see Carl E. Schneider, *Rights Discourse and Neonatal Euthanasia*, 76 Cal L Rev 151 (1988). For an extended critical treatment of American rights discourse, see Mary Ann Glendon, *Rights Talk* (Free Press, 1991).

discourse in America today, it should be no surprise that there are few areas of family law which someone has not suggested should be constitutionalized, and that there are not a few—entry into marriage, reproductive freedom, parental rights, and some aspects of child custody, for instance—which have been.

Rights solutions have seemed attractive in both areas of bioethics. *Roe v. Wade* obviously has something—often a great deal—to say about bioethical issues related to reproduction. And many of the questions about how medical-care decisions should be made have been analyzed in terms of rights, as the opinions in the *Cruzan* case suggest. Too many people (including me) have written too often and too long about *Roe v. Wade* to justify another inquiry into what it teaches about the usefulness of rights discourse. But we may learn something about three systematic problems with that discourse by looking more generally at family law's experience with it.

The first of these systematic problems is that "the origin, scope, justification, and purpose" of many of the constitutional rights at issue in family law cases are uncertain.[42] The rights at stake are generally what are loosely called "privacy" rights. These rights are essentially of recent origin, and the textual basis for them is slight. The case law through which they have been developed has not always labored to explore their nature or rationale. This is troubling on the familiar principle that in a democratic society courts, as non-majoritarian institutions, should not thwart decisions of majoritarian institutions without well-founded, well-articulated authority.[43] But it is also troubling at a more practical level. To see why, we need to understand something about how the Supreme Court analyzes family-law rights. The Court has commonly denominated most of these rights "fundamental," and it has (albeit somewhat erratically) imposed on any statute with which these rights conflict a generally unbearable burden of justification. Thus the question whether a litigant can

---

42. Schneider, 76 Cal L Rev at 158 (cited in note 41).

43. One of the unfortunate problems with rights discourse is that its underlying principles are—reasonably enough, under the circumstances—poorly understood by even quite sophisticated publics. It now seems to be true that a dismaying number of people simply expect the Supreme Court to put into law desirable social policy, not to interpret the Constitution. For a particularly sympathetic but still dismaying example of this confusion, see Pete Busalacchi, *How Can They?*, Hastings Center Rep 6 (Sept/Oct 1990).

assert a right is crucial. Yet because of its uncertainty about the nature of these rights, the Court has regularly had difficulty answering that question coherently and predictably.

The second systematic problem with family-law rights analysis lies in its difference from most other rights discourse. Ordinarily, we talk in terms of what I have called the Mill paradigm: "That is, we think in terms of the state's regulation of a person's actions. In such conflicts, we are predisposed to favor the person, out of respect for his moral autonomy and human dignity."[44] That predisposition also rests on our assumption that the state can bear any risks of an incorrect decision better than the individual can. "In family law, however, the Mill paradigm often breaks down, because in family law conflicts are often not between a person and the state but between one person and another person."[45] For example, we say that parents have a right to make decisions for their children. Yet we also say children have a right to life. If parents decide to deny their children treatment necessary to save their lives, how are we to choose between the two rights?

The third systematic problem with rights discourse in family law has been its inability to deal convincingly with the interests the state asserts to justify its infringement of rights. The Court often says that where a "fundamental" right is at stake, a statute must be "necessary" to serve a "compelling" state interest. But in practice the Court has been unwilling to apply this standard consistently or to define the standard's terms comprehensibly. In large part, this is probably because the test essentially requires the Court to compare two incommensurable values—the importance of the right with the importance of the state interest. On what scale, to take the example of *Zablocki v. Redhail*,[46] do you weigh the right of a person to marry against the state's interest in assuring that parents will support their children?[47]

---

44. Schneider, 76 Cal L Rev at 157 (cited in note 41).

45. Id.

46. 434 US 374 (1978). In *Zablocki*, the state had forbidden people to marry who already had children they could not or would not support. Id at 375.

47. For detailed criticisms of the Court's state-interest analysis in family-law cases, see Schneider, L & Comtemp Probs, at 79 (cited in note 24); and Carl E. Schneider, *State-Interest Analysis and the Channelling Function in Privacy Law*, in Stephen Gottlieb, ed, *Public Values in Constitutional Law* (Mich U Press, forthcoming 1993).

These three problems with rights discourse will often infect attempts to analyze bioethical issues in rights terms, if only because of the considerable overlap between family law and bioethics. Let us briefly examine some of the ways in which this happens by looking at the recent and familiar case of *Cruzan v. Director, Missouri Department of Health*.[48] The first of the problems of rights thinking we discussed was the obscurity of the origins, scope, justification, and purpose of family-law rights. I would suggest that such uncertainties about the nature of the right at stake in *Cruzan* explain much of the disagreement between the majority opinion and Justice Brennan's dissent. Justice Brennan vehemently insisted that Nancy Cruzan had a right to decide whether to receive food and water. But as the majority noted, "The difficulty with [that] claim is that in a sense it begs the question: an incompetent person is not able to make an informed and voluntary choice to exercise a hypothetical right to refuse treatment or any other right."[49]

The importance of Cruzan's incompetence lies in the nature of the right asserted for her. It is a right to decide. If we ask why we might attribute such a right to people, we are likely to suggest two reasons. The first is that people will make better decisions for themselves than the state can, since they know their own situation better than the state and have every incentive to make a more careful decision than the state would. But this rationale did not apply in *Cruzan*, since Nancy Cruzan could not know anything or respond to any incentives. The second reason we might attribute such rights to people is "out of respect for their status as independent moral agents."[50] But once again there is a problem with this basis for the privacy right in *Cruzan*, since "it makes little sense to attribute rights to people who cannot be independent moral agents."[51] The dissents needed, then, to explain why the origin

48. For a statement of the facts of *Cruzan*, see note 8. For an extended treatment of the case as a social, moral, and political question, see Carl E. Schneider, *Cruzan and the Constitutionalization of American Life*, 17 J Med & Phil (forthcoming 1992). For good analyses of *Cruzan* as a constitutional problem, see John A. Robertson, Cruzan *and the Constitutional Status of Nontreatment Decisions for Incompetent Patients*, 25 Ga L Rev 1139 (1991); and Yale Kamisar, *When is There a Constitutional "Right to Die"? When is There No Constitutional "Right to Live"?*, 25 Ga L Rev 1203 (1991).

49. *Cruzan*, 58 USLW at 4920.

50. Schneider, 76 Cal L Rev at 165 (cited in note 41).

51. Id.

and purpose of the right made it applicable to Cruzan. Instead, they formalistically and dogmatically insisted that she had a right to decide whatever her ability to claim, comprehend, or exercise it.

Cruzan instantiates the breakdown of the Mill paradigm in several related ways. First, that paradigm assumes a competent right-holder, which Nancy Cruzan was not. This might not have been crucial had she had only one right and had it been incontrovertible that she would have wanted to exercise it. But the second way in which *Cruzan* departs from the Mill paradigm is that she had not just one right, but almost a cacophony of rights, rights which potentially conflicted. She had a right to life; she had a right to refuse treatment necessary to preserve her life; she had a right (Justice Stevens believed) to have a decision made in her best interests. Finally, there were in her case two sets of potential right-holders—Nancy Cruzan and her parents. Yet the interests of the two sets potentially conflicted in the ways I earlier described.[52]

Finally, underlying much of the debate in *Cruzan* was the question how the state's interests should be analyzed. The dissents essentially argued that the state has no interest in the life of a person who does not want to live and that therefore the state cannot require that it be shown by clear and convincing evidence (instead of a mere preponderance of the evidence) that an incompetent person wishes to refuse treatment. This, I think, unduly (and characteristically) depreciates the state's interests. For one thing, the state has an interest arising out of its protective function. The evidentiary standard challenged in *Cruzan* was to be applied in *all* cases in which the issue was whether life-sustaining treatment should be denied an incompetent patient. It may well be that Nancy Cruzan would have wanted not to be treated in her circumstances. But I think it is at least constitutionally reasonable for the state to assume that, while many and perhaps most people would choose at some point to refuse medical treatment, most people prefer life to death and will struggle to retain it as long as

52. One way of resolving the potential conflict would be to say that Cruzan's parents had no distinct rights of their own, but were merely exercising her rights for her. However, this does not really make the problem go away, since there remain not only the questions whether her parents in fact had no right of their own and whether a right like Cruzan's could be exercised by an unappointed proxy but also the fact that her interests and their interests potentially conflicted.

they feel they can. On this view, the state protects people who cannot protect themselves by setting a general evidentiary standard that errs on the side of treatment.[53] This view is made more plausible by the consideration that the people—the family—who will usually be seeking to end treatment will be people who will not uncommonly stand to benefit in some way from doing so. And even when, as will ordinarily be the case, patients ultimately do not need to be protected from their families, the state can point out that they may still need to be protected from the other people who may participate in making decisions about the patient—namely, the relevant medical personnel.[54]

---

53. It is, as the dissents pointed out, no doubt true that Missouri's evidentiary standard would sometimes result in treatment being given where the patient would not have wanted it. See, for example, *Cruzan*, 58 USLW at 4926 (Brennan dissenting). But as the majority noted, such a result is quite unremarkable in our legal system. Id at 4921. We regularly decline to give effect even to a clearly expressed intention where that intention has not been given the proper legal form, as the laws of wills, gifts, conveyancing, and contracts (to name only a few) all testify. We do so in part for reasons of efficiency: Where people have followed the correct legal forms in expressing their wishes, we are relieved of the burden of ad hoc inquiries into their true intent. But we also do so because we have the deepest doubts about the success of any such inquiries. In addition, we impose on people the obligation of making their preferences clear so that everyone who needs to know can know with confidence what those preferences are. Finally, we ordinarily decline to enforce preferences that are not expressed in the correct legal form because of our fear that even a clearly and accurately expressed wish may not be what the person truly wants. We have all had the experience of thinking that if some situation arose we would want some particular result, but nevertheless discovering that, when pressed to make an actual decision, our impulse was not our true wish. For a moving expression of such a discovery, see Vicki Williams, *The Horror Is Worth It*, Newsweek 14 (Oct 9, 1989), a wife's account of how her terminally ill husband reversed his initial decision to refuse aggressive treatment. Less dramatically, see Jay J.J. Christensen-Szalanski, *Discount Functions and the Measurement of Patients' Values: Women's Decisions During Childbirth*, 4 Med Decision Making 47 (1984), which reports that a significant proportion of women who had chosen to forego analgesics during childbirth changed their minds during delivery. (It should be said that not all studies of patient preferences indicate this kind of instability. See, for instance, Maria A. Everhart and Robert A. Pearlman, *Stability of Patient Preferences Regarding Life-Sustaining Treatments*, 97 Chest 159 (1990).) The forms and formalities associated with preparing and signing legal documents are intended to bring home to their signers that a binding and consequential decision is being made and thus to promote as "true" a decision as possible.

Part of the problem in *Cruzan* is probably that Nancy Cruzan was caught in a transitional period when new legal responses to the problems of incompetent patients are being created and publicized. It is possible that, as living wills and durable powers of attorney become more common, people like Cruzan will come to know about them and, where they want to, sign them. At least at that point it will be more reasonable to expect people to do so and to deny effect to any wishes they express that are not in a form clearly announcing that their wishes are intended to have legal effect.

54. On the propensity of some physicians to see treatment issues as exclusively

In addition, is it true that society has no interest in the lives of its citizens once they have decided not to live? Suppose, for instance, that A had irrevocably decided to commit suicide and had taken poison which would inevitably result in his death. Suppose further that B then killed A. Is B innocent of homicide because A's life has ceased to be of interest to society? Surely not. B is guilty of homicide partly because of the social interest in maintaining a sense of the sanctity of human life in order to encourage people to respect it. But the social interest in A's life also arises out of the belief that few things are more basically important than human life, that it is valuable in itself and not just to the holder, that "each man's death diminishes me." If there is a social interest in rocks, louseworts, and snail darters, why not in people's lives?[55]

In criticizing contemporary rights discourse and in surveying its limitations, I have not intended to say that rights solutions should never be sought, that that discourse does not serve valuable purposes, or that all the problems with our rights discourse are insuperable. But in America today rights solutions have so powerful an appeal that the greater danger is that they will be unreflectively adopted and dogmatically defended. Thus I have been more concerned here with some of the cautionary experiences family law has encountered in using rights discourse than with the well-known advantages that flow from it.

VII. THE FACILITATIVE FUNCTION

Another response to the standards problem has recently grown abundantly in popularity. This response is to expand what I call the law's "facilitative function." The facilitative function

---

medical and not at all moral or social, see Allen E. Buchanan, *Medical Paternalism or Legal Imperialism: Not the Only Alternatives for Handling* Saikewicz*-Type Cases*, 5 Am J L & Med 97 (1979). For a more realistic look than any of the opinions in *Cruzan* offers of the actual practice of "informed consent," see B.W. Levin, *The Culture and Politics of "Baby Doe" Decisions*, Paper presented at the 108th Annual Spring Meeting of the American Ethnological Society (1986); and Zussman, *Intensive Care* (cited in note 13).

55. Justice Scalia's concurring opinion took an even stronger line on the state-interest question. It argued that what was at stake in the case was whether Cruzan could commit suicide and that the state's interest in preventing suicide was solidly established by centuries of statutory and case-law precedent. *Cruzan*, 58 USLW at 4924-26 (Scalia concurring).

allows people to deploy the law's power to arrange their lives in ways they prefer. It has two primary forms. First, it offers people a legally enforceable way of specifying how their affairs should be handled, as when it allows them to specify in a will how their assets will be distributed. Second, it provides people with a legally enforceable way of arranging their relationships, as when it allows them to enter into contracts. An attraction of both forms is that they permit people to choose their own standards, so that the law need not prescribe standards for them.[56]

Family law has long resisted the use of contracts in most family settings. Recently, however, it has become markedly more willing to allow couples both before and during marriage to enter into contracts regulating some of their relations during and after marriage. And it has also become more welcoming of contracts between unmarried cohabitants. Nevertheless, family law's attitude toward contracts remains cautious. Why? For one thing, family relationships often involve emotive, fluid, and personal attitudes and behavior that are not consonant with the kinds of rationalistic and calculating attitudes that we associate with contract law. For another thing, people in family contexts may be unable or unwilling to bargain aggressively, to guard against internal and external emotional pressures, and to foresee far into the future how they will feel about complex and intractable problems they cannot now even imagine. Family law's protective function thus may well call for it to safeguard at least the weaker party to the contract. Further, many contracts affect people besides the contracting parties. These people will not have had the choice of standards the parties to the contract had, and so for them the facilitative function will not have solved the standards problem. What is worse, these third parties may be injured by the contract, thus calling the protective function into play. And, of course, there is always the awkward fact that many people will not make a contract despite every incentive to do so, just as many people will not write a will despite centuries of encouragement and admonishment. In sum, family law has resisted contract as an

56. For an exploration of the facilitative function and of contract in family law, see Schneider, *Family Law: Cases and Materials* (cited in note 7). For a more extended and favorable view of the use of contract in family law than I offer in this essay, see Marjorie M. Shultz, *Contractual Ordering of Marriage: A New Model for State Policy*, 70 Cal L Rev 204, 244-65 (1982).

ordering principle because of doubts about the appropriateness of using a commercial mechanism in a personal setting, and families seem often to share that resistance.

Similar concerns will inhibit solving bioethical problems through the facilitative function. Consider, for instance, surrogate-mother contracts. Can prospective surrogates think with the rationality contracts require about something as fiercely "emotive, fluid, and personal" as how they will feel about giving up the child they have borne for someone else? Will they be economically or psychologically vulnerable and be pressured into making agreements they will ultimately regret and even abjure? Will they be willing and able to bargain aggressively to protect their interests? Will they foresee when they sign the contract how they will feel when the time comes to execute it? And, of course, the contract produces a person who was not party to it—the child. The law will have some interest in protecting that child, and in doing so the law will again be returned to the standards problem.[57]

The other aspect of the facilitative function—the one which allows people to recruit the law's power to effectuate their individual intentions—has also emerged as a solution to some bioethical problems. Primary examples are the living will and the durable power of attorney. But these devices are also subject to some of the uncertainties that characterize the facilitative function. How far will people signing such a document fully understand the circumstances in which it might be applied, fully have thought about their own feelings about these distressing subjects in the present, and accurately anticipate how they will feel in the future? Will they have been unduly influenced by the people around them? These questions are intended to suggest, of course, that medical decisions are brutally hard to make under the best of circumstances and that making them in a present that might be unrecognizably different from the future is to make them under quite deplorable circumstances. It is thus not surprising

57. For a fuller treatment of this problem, see Carl E. Schneider, *Surrogate Motherhood from the Perspective of Family Law*, 13 Harv J L & Pub Pol 125 (1990); Martha A. Field, *Surrogate Motherhood: The Legal and Human Issues* (Harv U Press, 1990); Symposium, In re Baby M, 76 Georgetown L J 1719 (1988). For discussions of the related question of a market for adoptive children, see Elisabeth M. Landes and Richard A. Posner, *The Economics of the Baby Shortage*, 7 J Legal Stud 323 (1978); Symposium, *Adoption and Market Theory*, 67 BU L Rev 59 (1987).

that there is evidence that advance directives have a good deal less effect than we might like to believe.[58]

All these considerations suggest that we should constrain our ever-soaring hopes that the devices of the facilitative function will solve the bioethical dilemmas we now face. They also counsel us that, if the facilitative function is to be given substantial legal standing in bioethical decisions, it should at least be in a carefully formalized way. Casual substitutes for careful thought should not be encouraged, for the facilitative function achieves its deepest justification only when it backs with the force of law people's genuinely considered wishes. The questions I asked above are hard enough where the prospect of signing a binding legal document has brought home the fact that serious issues are being resolved.[59] These questions become next to impossible when, as in *Cruzan*, the only evidence is the possibly quite casual remarks to friends of a young person who is not aware that what she says will have actual consequences and who does not expect to have desperate medical problems for decades. And when the patient was never in his life competent to formulate an opinion on treatment, any attempt to decipher his intention must be wholly fictional.[60]

VIII. CONCLUSION

This attempt to glean lessons for bioethics from family law has yielded no determinate answers or easy principles. I have suggested that family law has recently struggled to avoid the

---

58. See, for example, David Orentlicher, *The Illusion of Patient Choice in End-of-Life Decisions*, 267 JAMA 2101 (1992). Lawyers are regularly surprised when the world ignores legal rules, but by now they should not be. See, for example, Stewart Macaulay's classic study of the use of contracts in business: *Non-Contractual Relations in Business: A Preliminary Study*, 28 Am Soc Rev 55 (1963). On the distance between law and life in one significant area of family law, see Schneider, 1991 BYU L Rev at 203-209 (cited in note 2).

59. For a masterly demonstration of just how baffling those questions can be even in optimal conditions, see Patricia D. White, *Appointing a Proxy Under the Best of Circumstances*, 1992 Utah L Rev 849.

60. See, for example, *Superintendent of Belchertown State School v. Saikewicz*, 370 NE2d 417 (Mass 1977), where the court struggled hopelessly to solve the problem by attempting to do what the patient would have done had he been competent to decide for himself. Id at 431. For a good statement of the limits of this "substituted judgment" procedure, see Allen E. Buchanan and Dan W. Brock, *Deciding for Others: The Ethics of Surrogate Decision Making* 112-22 (Cambridge U Press, 1989).

standards problems. Yet I have argued that each method of doing so is itself importantly flawed and sharply limited. I must confess that, if anything, this survey has been too pessimistic, that it has looked more assiduously for the drawbacks than the benefits of each approach. I should also say that my survey has confined itself to examining broad approaches, rather than seeking the surely valuable lessons to be learned from family law's concrete, common-law resolutions of particular fact-patterns.

On the other hand, if my cautionary approach is essentially correct, two conclusions might follow. The first is that the standards problem might be confronted directly. My own inclination is that there is something in this. I am not persuaded, despite some real evidence to the contrary, that the processes of democratic government are wholly incapable of resolving the kinds of value conflicts that family law and bioethics present or that allowing them to do so is wholly incompatible with a free society. But I readily admit that this is only an inclination, and that I have not fully worked out all my own views on the subject.[61]

The other conclusion that might be drawn from family law's quandary is that we must content ourselves with imperfect solutions to the perplexing issues family law and bioethics present us. Perhaps one reason the approaches I have canvassed seem so inadequate is that too much has been expected of the law. We want the law always to reach the right result, and when in a given case it fails to do so, we demand in our distress that somehow the law should be changed. We insist that the law must get every case right, and we are indignant when it fails (by our lights) to do so. But family law and bioethics, like much of law, are areas where it is hard to know what the right result is, where often there will be no right result, and where there is no way of assuring that the right result will consistently be reached. Both areas, like much of law and life, involve what have been called "tragic choices"—irreducible conflicts between legitimate interests.[62] We may, then, have to content ourselves with picking

61. For an attempt to work out some of those views in the particular context of the law of alimony, see Schneider, 1991 BYU L Rev (cited in note 2).

62. For a general description of this problem, see Guido Calabresi and Philip Bobbitt, *Tragic Choices* (WW Norton & Co, 1978). For a further treatment of it, one that has much to say in particular about its appearance in the dispute over abortion, see Guido Calabresi, *Ideals, Beliefs, Attitudes, and the Law: Private Law Perspectives on a Public*

eclectically whichever approach seems best adapted to the particular problem at hand, consoling ourselves with the realization that often human institutions can do no better than to muddle through.

Law Problem (Syracuse U Press, 1985).

Appointing a Proxy Under the Best of Circumstances

Patricia D. White[*]

We are commonly told that if we wish to ensure that our end is not unduly prolonged by extraordinary medical treatment we should either execute a living will or appoint someone to be our proxy medical treatment decision maker. Estate planners, medical personnel, and now even the federal government[1] encourage us to choose a surrogate to act in our stead in case of our own incompetence. If we do this, it is said, we can avoid becoming a victim of the debate about who should decide when to terminate an incompetent patient's life support systems if the patient, when competent, never articulated a preference nor appointed a proxy decision maker.

The debate itself is an arresting one. It has all the elements of high drama and forces us to wrestle with a set of moral problems about which many feel strongly but which, like all such problems, is probably irresoluble outside the context of a deeper and pervasive moral theory. The debate becomes more urgent for us, however, because the realities of modern medicine require that it be resolved, at least as an issue of public policy. And if the question of who should decide is not difficult enough, we also struggle with the difficult question of what standard or standards should guide the decision maker. Should the standard be the best interest of the patient as conceived when the patient was competent; the best interest of the patient now; the best interest of the decision maker; the best interest of the state; or is it the decision which best approximates the decision which the patient—if competent—would now make knowing what the decision maker knows about the patient's present condition and prognosis? As this highly-charged discussion evolves—and as the advice to appoint a proxy decision maker is increasingly given—I am struck by how little systematic attention has been given to a set of issues which

---

* Visiting Professor of Law, University of Michigan.

1. *See, e.g.*, Patient Self Determination Act, Pub. L. No. 101-508, §§ 4206, 4751, 104 Stat. 1388, 1388-115, 1388-204 (1990) (codified at 42 U.S.C.A. §§ 1395cc, 1396a (West 1992)) (directing health care providers to advise patients of their rights under state law to execute advance directives).

is both conceptually prior to these questions and of great practical significance. We have not stopped to look carefully at just what possible choices *competent* people might make when they do in fact appoint proxy medical decision makers. I have in mind a set of issues different from the issues surrounding the content of the patient's living will.[2] This paper analyzes the choices available to a competent person when appointing a proxy. My hope is that if we can clarify these issues, we can begin to help people articulate their own choices more clearly and have a more satisfactory discussion of what is at stake in a setting where no decision maker has been appointed for an incompetent patient.

A competent person might appoint a proxy decision maker at a time when he was healthy or at a time when he was sick. It seems probable that one's health at the time of appointing a proxy decision maker could influence significantly what one intends by the appointment. Even self-aware people find it difficult to predict accurately how they would react to some hypothetical future crisis. The perceived immediacy of appointing a proxy decision maker would clearly make the choice made by a competent but extremely ill person different from the choice made by the same person when he enjoyed perfect health. But while it seems clear that a person's health might well affect the intent of his proxy appointment, it is not at all clear what that effect would be. For example, some gravely ill people might be able to see and articulate their desires more clearly than they could while healthy, while others might find themselves less able to determine what they would want. This lessened ability might in no way be indicative of incompetence; rather, it might be a function of denial, fear, ambivalence, genuine indecision, or some combination of these or other reactions to the situation. Unfortunately, then, we have a problem with our analysis at the outset because we cannot even say with confidence that the last proxy executed by a competent person most accurately reflects his wishes. This comment may seem odd, since it might appear that even a gravely ill competent patient who is in a state of denying his likely imminent need for a proxy decision maker *is* reflecting his latest thoughts on the subject when he appoints a proxy a few days before lapsing into unconsciousness. But while he is perhaps reflecting his latest thoughts on the subject he may

2. This is a distinct area which is not necessarily well thought out either.

well not be reflecting his considered thoughts on the subject. Those might have been reflected in a prior inconsistent proxy directive. It might well be that had the patient remained competent and gravely ill for a long enough time, he might have come to terms with his situation and ratified his first proxy designation or perhaps executed one altogether different from either of the other two.

Putting these complications aside for the moment, assume that a rational, healthy person is asked to think hard about appointing a proxy medical decision maker. Assume, too, that this person is not being asked to write a living will or indicate what particular treatments he would like his proxy decision maker to choose for him in particular circumstances.

There are at least three rather different sorts of situations in which a person might find himself needing a proxy medical decision maker. The obvious and much discussed case is the one in which the patient is either unconscious or otherwise incompetent and is in quite dire straits, unlikely ever to recover. The second is the circumstance where the patient is temporarily unconscious or otherwise incompetent but is expected to recover fully. Somewhere between these two extremes on the inevitable continuum is the case where the unconscious or otherwise incompetent patient might or might not recover. There are important differences between these circumstances from the perspective of a person thinking about appointing a surrogate medical decision maker. Indeed, the rational person might well look for different things from the decision maker appointed in each of these three circumstances.

If, for example, one wanted to provide for the possibility that unanticipated medical decisions might need to be made in connection with ongoing surgery, during which she would be unconscious but from which she was fully expected to recover, some might request that the decision maker's standard be "the patient's best interest," while others might require that the decision maker adhere to a substituted judgment standard. These two standards might be substantially different in this circumstance, and the person best suited to apply each standard might be different. If, as hypothesized, full recovery were the likely outcome, the patient might reasonably appoint her physician as her proxy decision maker with the instruction that the doctor act

in the patient's best interest. The patient in this situation would likely want those medical choices made which would maximize her chances for a medically optimal outcome. If a person had selected a physician in whom she had confidence, she would presumably have confidence in that physician's judgment throughout the episode. On the other hand, the patient might hate pain or may be particularly squeamish about gruesome-sounding procedures. Such a person might not be able to count on herself to make those medical decisions which are in her own best interest. We all know people like this, and probably nearly all of us are more like this than our rational selves care to admit. This is the person who puts off going to the doctor when she has some disconcerting symptom. This is the person who refuses to exercise when he is told that his weight is dangerously high. This person cannot count on himself to act in his own best interest when he is making his own medical choices. If his proxy decision maker were instructed to use a substituted judgment standard when making decisions for the unconscious patient, and if that instruction and standard were taken seriously, then the proxy might make choices that are not in the patient's best interest.

If our competent person, seeking to appoint a proxy, wanted someone who understood and would honor her aversions, she would clearly choose someone other than her physician, presumably someone to whom she was very close. She would instruct her proxy decision maker to make the same choices that she herself would have made. In other words, the choice that the physician would make might differ from the choice that the patient herself would make. Although it might seem misleading to say that the former decision is made in the patient's best interest while the latter, the one the patient would have made, is not in the patient's best interest, this is not necessarily true. People not only sometimes do act against their own best interest, but they sometimes want to act against their own best interest. Wanting to act in a particular way does not change the nature of the act into one which is in the person's best interest. This is certainly true if the best interest standard is thought to apply to the person's medical condition and the substituted judgment standard is thought to represent the choice which the patient himself would actually have made had he been able to make it. So even in the simplest of cases, the person seeking to appoint a medical decision proxy is

faced with a nontrivial choice: Should he appoint someone who will act in his best interest or should he appoint someone who will act as he would act, in accordance with the substituted judgment standard? Not only is he faced with choosing a standard but each standard might suggest a different proxy.

Things become considerably more complicated as the nature of the circumstance in which the proxy is to act changes. If you are seeking to appoint a proxy to act as your surrogate in a circumstance where your recovery is so unlikely as to be essentially impossible, there are other possible standards which you might wish to have applied to the surrogate's decisions. Of course you might request that the surrogate act in your own best interest or act as you would. Alternatively, you might believe that except for the management of your pain, you would no longer have a stake in the details of your treatment, since its outcome is certain ultimately to be death. You might therefore want to choose as a surrogate someone who would have something at stake; perhaps the person who would be financially responsible for your care, or perhaps someone who would be emotionally affected by your death. You might in fact want to choose as a surrogate the person who had the most at stake in your treatment and instruct that person to make decisions about your medical treatment in the surrogate's best interest. This is not something that everyone would wish to do, nor is it something that everyone ought to do, but surely it is something that a reasonable proxy appointer might do.

We have not yet exhausted the possibilities. As indicated above, you might believe that you would no longer have a stake in the outcome and that others would. However, instead of selecting a proxy whose own stake is great, you might choose instead a person whom you believe is best able to consider and weigh the interests of everyone who has something at stake. This might lead you to appoint someone who personally has little or nothing at stake, but whose judgment you trust. Since, as I have described this situation, this judgment would not primarily be medical, this person might be some relatively neutral party, such as a clergyman or other close family advisor, a trusted friend, or a social worker. Or you might choose someone who had something significant at stake, but whom you nonetheless asked to give her own interests only the weight they deserved in comparison with

others' interests. Finally, you might want to instruct your surrogate to act as he wants to. You don't give him guidance, you don't instruct him to act in his own best interest, and you don't instruct him to act in the best interest of everyone with something at stake. Presumably, the distinction between what the surrogate wants and what is in the surrogate's best interest is as real as the distinction between what you want and what is in your best interest.

But the story does not end here. You might believe that you *would* have an interest in the outcome, even if you were unconscious and without any prospect of recovery. You might believe that your interest was only one of the interests to be considered, or you might believe that your interest in the outcome was paramount. If you believed that your interest was but one of the relevant interests, you might want a surrogate who would weigh all of the interests, including yours, in making choices on your behalf. Once again, you might choose a neutral but trusted person for this task. On the other hand, if you believed both that your interest in the outcome of your treatment would not be extinguished until your death, and that your interest overrides all other interests, you would select as a surrogate either a person who would choose as you would choose or a person who would use your best interest as his standard and whose understanding of what constitutes your best interest is consistent with your own. If your overriding interest is your life, you would probably choose a different surrogate from the one you would choose if your overriding interest is "dying with dignity." And you might make still a different choice if your primary concern is that you not burden your family or loved ones. Ironically, if your primary concern were the last one, you might instruct your surrogate to make choices on the basis of what is in the best interest of everyone who would be affected by the decision, other than you. You would do this precisely because you would like to effectuate what you consider to be your overriding interest.

The discussion so far has centered on healthy, competent people trying to determine whom they would appoint as proxy medical decision makers, first where unanticipated medical decisions might need to be made during a period of temporary unconsciousness or incompetence but where the patient's full recovery is expected, and second in the radically different circum-

stance where treatment decisions would need to be made for a patient who is almost certainly not going to recover. Unfortunately, the need for a proxy decision maker is probably most likely to arise in a situation which is more ambiguous than either of these extremes.

The ambiguity often arises because it is difficult to predict exactly what degree of recovery is likely. If it is difficult for a physician to predict reliably the course of a critically ill patient's treatment and recovery, it is even more difficult for a proxy decision maker to apply his assigned standard for decision making. If the decision maker has been instructed to act in the patient's best interest, the decision maker is presented with an impossible task, because the factual data necessary to determine the patient's best interest are patently inadequate. If, as a rational, healthy person, you wanted to appoint a surrogate medical decision maker to act for you in this ambiguous circumstance, and if you wanted the surrogate to act in your best interest, then you might place more importance on the surrogate's ability to achieve a relatively sophisticated understanding of your medical situation and prognosis than you would in the situation where your prognosis is unambiguously bleak. This might, of course, argue for choosing your physician as a proxy, but not necessarily. If you had wanted the proxy to apply the substituted judgment standard, you might have chosen a close friend or a relative in the earlier less ambiguous situations. Those same reasons apply here. You might want a proxy who knows you and your values well enough to allow him to make an informed judgment as to what is in your best interest. The proxy's determination might still be different from the determination you would make yourself, but your own ability to deal with a life of diminished capacity is surely relevant to a determination of your best interest. Whereas determining your best interest when full recovery is clear seems to be largely a medical judgment, and whereas determining your best interest when failure to recover is clear seems to be largely a judgment about you and your values, determining what is in your best interest in the ambiguous situation seems to require a proxy who can combine both judgments in making a much more complicated determination. It is entirely plausible to imagine, therefore, that a rational healthy person might select a different proxy for each

of the three circumstances, even though the standard governing each proxy is the patient's best interest.

The judgment that the proxy is being asked to make is correspondingly complicated if the rational, healthy proxy appointer selects any of the other possible standards. As indicated above, a healthy appointer might select as a surrogate the family member with the most at stake if the appointer were in a persistent vegetative state with no hope of recovery, and might instruct that proxy to act in the proxy's own best interest. Nonetheless, the appointer could conceivably not want the proxy to be placed in the position of determining the appointer's fate in a medically complex or uncertain situation, even though the appointer wants the standard for decisions to remain the best interest of the proxy. The appointer might recognize that the required decisions could be so complex that the proxy would simply not be the person best able to make the determinations which would best protect the proxy's own interests.

The medically ambiguous situation suggests yet another type of proxy arrangement which might be selected by the rational healthy person in anticipation of finding herself in this circumstance. She might believe that her own best interest should be paramount until, and only until, the interests of others outweighed it. This point could be reached when her level of probable recovery were sufficiently low or when the burden on others of maintaining her had become sufficiently high. At this point the proxy appointer might want to shift proxies, shift standards to be applied by the proxy, or both. Indeed, in medically ambiguous situations the rational person might want to establish a mechanism whereby someone is appointed whose sole job is to determine when this point has been reached—mediating, as you will, between different proxies, different standards, or both.

I have described the need for a proxy as falling on a continuum ranging from the patient's short-term need for a surrogate decision maker during a period of temporary incompetence or unconsciousness pending expected full recovery, to the situation exemplified by the patient in the persistent vegetative state with no hope of recovery. In anticipation of finding himself at some time at some point along this continuum, the rational healthy person might choose different proxies, different standards, or both for the proxy to apply at various points on this continuum. As the

preceding discussion indicates, the process of appointing a proxy medical decision maker, if approached seriously, ought to be a far more complicated one than current practice would suggest. The standard forms of durable powers of attorney for health care decisions do not begin to reflect anything of the subtlety which the question they seek to answer demands.

A durable power of attorney for health care decisions is often accompanied by a living will or advance directive. An advance directive typically specifies the conditions under which the person executing it would not wish to be kept alive by various, often specified, forms of so-called heroic measures. It does not indicate who is to make medical decisions on the patient's behalf; rather, it purports to be a document whereby the patient makes his own decisions with respect to a subset of potential future circumstances. This distinction is important because it underscores the very great practical limitations which are built into the notion of the living will. Whereas the impetus for honoring the instructions set forth in an advance directive arises from the presumption that the document expresses the patient's own wishes and thus represents his autonomous choice, it is a choice which the patient has necessarily made on the basis of incomplete information.

A living will is a very crude instrument to use for making actual medical decisions. It is far more effective as a device that allows a person to make known his attitude about a whole range of issues surrounding medical intervention at or near the end of life. A living will is best viewed as setting forth standards for decision making which should be taken into account by whoever actually makes the decisions governing the incompetent patient's medical care. The reason that an advance directive, under normal circumstances, should only be regarded as advisory rather than determinative is that all it can express is what a competent person thought she would want were she to become incompetent and be in a situation generically like the one she turns out actually to be in. This kind of preference expression is clearly different from the sort of informed consent that we require of competent patients before undertaking medically significant procedures or treatment. The difference suggests why it is a mistake to conceive of an advance directive as expressing an incompetent patient's autonomous choice in any specific circumstance. We need not doubt that the patient really meant what she said nor that what she said

reflected her attitudes and values. But we must acknowledge that, typically at least, the patient did not in her advance directive address the precise question which needs to be answered; namely, how, at this time and in these particular circumstances, she should be treated. Thus while we might wish to argue that respect for a patient's autonomy requires paying attention to the values and attitudes expressed in his advance directive, it does not make sense to justify relying on the instructions in an advance directive on the grounds that they express the patient's decision about particular medical treatment.

It is worth looking briefly at the case of the advance directive or living will because the same observation can be made about a competent person's appointment of a proxy medical decision maker. The assumption behind honoring such an appointment at the time of a patient's incompetence is that the appointment represents the patient's wishes, and we justify honoring his wishes by suggesting that they represent his autonomous choice. But to the extent that we have reasonable doubt about what his actual wishes were, then we seem less bound to try to honor them. My analysis of the considerations that a well-thought-out proxy appointment might entail suggests that in most cases where a medical treatment decision proxy has been appointed, the appointing person will either not have been asked or not adequately focused upon the questions that he needs to answer in order to know what he really wants. If he did not answer the right questions then it is difficult to justify honoring what he said on grounds of autonomy. As a practical matter, our efforts as lawyers, philosophers, and physicians should be directed at helping proxy appointers frame the questions whose answers are necessary to gain a proper understanding of the appointer's intentions. The analysis of this paper is intended as a first step toward that goal.

In the absence of sufficiently helpful directions from a competent person about who should speak for him during his incompetence and in accordance with what standard, we need a default decision maker or proxy. Typically, of course, physicians and nurses consult with a patient's family and try to reach a consensus about how best to act. In the vast majority of cases this sort of consultative procedure works well. But problems inevitably arise when there is dissension between the doctors and the family,

among the family members, or when, as in the *Cruzan*[3] case, the hospital administration seeks judicial approval of its action in order to protect itself from legal liability for a decision which could be construed as hastening a patient's death. This default mechanism could take various forms. It could either create a presumption in favor of a decision maker or process by which decisions will be made, or it could presume that certain standards will govern all medical decisions. Thus, for example, the default position might favor maintaining a patient's life at whatever cost and at whatever diminished capacity.[4] On the other hand, it could presume a decision maker rather than a result. For example, the presumption could specify certain family members in some order of preference.[5]

Mistakes would doubtlessly be made under any presumption—at least if the standard for determining what constitutes a mistake is what that patient would have done when competent were he to observe his own plight as an incompetent patient. Although it is difficult really to argue for this view, my own sense is that more mistakes would probably be made if a result were presumed than would be made if the presumption were instead in favor of family decision makers. The available data seems to indicate that most people would not prefer life at all costs, although some would.[6] An overriding principle in favor of prolonging life, therefore, would clearly result in some lives being extended in ways the patients themselves would not have chosen.

---

3. Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2846 (1990).

4. Yale Kamisar takes something close to this view. He argues that to allow others to weigh the costs and benefits of the patient's existence is to embark on a very dangerous slippery slope. *See* Yale Kamisar, *When Is There a Constitutional "Right to Die"? When Is There No Constitutional "Right to Live"?*, 25 GA. L. REV. 1203, 1203 (1991); Yale Kamisar, *Some Non-Religious Views Against Proposed "Mercy-Killing" Legislation*, 42 MINN. L. REV. 969, 1030-41 (1958). This was also the effect of the position adopted by Missouri and its Supreme Court in Cruzan v. Harmon, 760 S.W.2d 408, 424-27 (Mo. 1988) (en banc), *aff'd*, Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2845 (1990).

5. Nancy Rhoden argued that there should be a presumption in favor of families as decision makers. Nancy K. Rhoden, *Litigating Life and Death*, 102 HARV. L. REV. 375, 390-94 (1988). Several states have statutes which specify an ordered list of family members as surrogate decision makers. *See, e.g.*, UTAH CODE ANN. § 75-2-1105(2)(b) (Supp. 1992).

6. One small study found that elderly patients have both an expectation of and a preference for familial decision making in the event of their future inability to make their own medical decisions. Dallas M. High, *All in the Family: Extended Autonomy and Expectations in Surrogate Health Care Decision-Making*, 28 GERONTOLOGIST 46, 48-49 (Supp. 1988).

Although it is hard to conceive that a court would articulate an overriding principle in favor of removing life support systems, such a principle would certainly yield a result different from what people like Helga Wanglie[7] would choose. By contrast, a presumption in favor of a specified family decision maker would at least allow for the possibility that different decisions would be made for different patients, and thus acknowledge the fact that people's preferences, as expressed when they are competent, differ.

A thorough analysis of potential default mechanisms would be a complex and difficult undertaking. Such an undertaking is not the project of this paper. The point here is to begin an analysis which properly precedes the choice of a default mechanism. Until we understand the range of choices which a healthy competent person might rationally make when appointing a proxy for medical decisions, we cannot make much progress helping others to articulate their choices. The Patient Self Determination Act[8] seeks to increase the use of advance directives and proxy appointments by hospital patients. This goal reflects a sense that these documents reliably represent a person's wishes. My analysis, however, suggests that the typical proxy appointer will not have been asked or have focused on the questions whose answers are necessary for this to be true. There is real work to be done.

7. *See In re* Wanglie, No. PX-91-283 (Minn. Dist. Ct. June 28, 1991). For a discussion of the *Wanglie* decision, see Cathaleen A. Roach, *Paradox and Pandora's Box: The Tragedy of Current Right-to-Die Jurisprudence* 25 U. MICH. J.L. REF. 133 (1991).

8. Pub. L. No. 101-508, §§ 4206, 4751, 104 Stat. 1388, 1388-115, 1388-204 (1990) (codified at 42 U.S.C.A. §§ 1395cc, 1396a (West 1992)). For a brief discussion of the Act's rationale see Elizabeth McCloskey, *Between Isolation and Intrusion: The Patient Self-Determination Act*, 19 LAW, MED. & HEALTH CARE 80 (1991).

The Roles of the Family in Making Health Care Decisions for Incompetent Patients

Leslie P. Francis[*]

I. INTRODUCTION

This Article is about the roles of the family in making health care decisions for incompetent patients. It argues that complex moral reasons call for the participation of families in decision making for incompetents. However, these moral reasons do not support a single model of the family's role for all incompetents. Rather, they suggest important differences among the roles family members should play in decision making for different kinds of incompetent patients: formerly competent adults, never competent adults, or infants and children.

Briefly summarized, the differences in roles are as follows: First, in making decisions for formerly competent adults, the family's principal role should be to help with the recognition of the patient as autonomous chooser, insofar as that can be achieved. Thus, the family may be called upon to provide information about the patient, such as the patient's expressed preferences about health care decisions (when they are available), values, and approaches to making decisions. Second, in making decisions for never competent adults, the family's principal role should be that of concerned advocate for the interests of the patient. Here, the family attempts to ensure that the patient's interests are understood and considered. Third, in making decisions for infants and young children, the family's role is to construct, as well as to pursue, a reasonable account of the child's best interests. These differences among roles are important, and developing them will be the task of the first two parts of this Article.

The third part of this Article will consider how families have been legally included in—or banished from—health care decision making for the same groups of incompetent patients. Recent American law—at least since the potentialities of modern medical

* Professor of Law, Associate Professor of Philosophy, University of Utah.

care have come under judicial scrutiny—has confronted a remarkable range of situations in which the roles of families have been at issue. Unfortunately, however, the law has been more than a little unclear about how to treat families. Trends, of course, are always difficult to summarize, but there have been notable instances in which families have been allowed a great deal of latitude in decision making for formerly competent and never competent adult patients. However, families have been given little say in decision making for young children and particularly for infants. The law, in short, may be morally backwards.

Several introductory comments are important to avoid misunderstanding the argument which follows. First, there are well-known problems with family decision making for incompetents, which are largely bypassed in this discussion. Families may have conflicts of interest. The most obvious conflicts are money and time; the continuing need for care may drain family finances or become a constant burden. But there are other kinds of conflicts as well, for example, emotional stress: families may prefer the patient's death to the daily renewal of grief as a disabled relative continues to survive. Even if these conflicts of interest are relatively minimal, families may have difficulty shouldering the responsibilities of decision making. It may seem unfair—a genuine abdication—for health care providers to expect families to take responsibility for difficult decisions. Shifting the burdens of decision making to the family may seem particularly problematic if families are coming to terms with the sudden shock of illness or disability. It may seem downright cruel if family members have themselves been injured or physically affected by the events of the patient's illness or injury, such as luckier (and possibly guilt-ridden) survivors of a devastating accident or a mother who has just given birth to a compromised infant. Some of these concerns—time, money, or emotional stress—are likely to be present when families are involved in the care of any of these patients. Other concerns—the shock of confronting an unexpectedly disabled newborn, together with the mother's own physical state—may apply differently to different types of patients and in different types of circumstances. Although this discussion does not focus on the general problems with family decision making, they should not be forgotten. Particularly severe risks of conflicts of interest, for

example, may in certain cases justify different legal treatment of the family's role.

Second, this discussion will largely bypass issues in defining the family. Is "the family" defined by biological relationships and, if so, which ones? Is it defined by legally recognized relationships, including some biological relationships but also including adoption and marriage? Can individuals construct their own familial relationships, such as through surrogacy contracts or same sex marriages?[1] State statutes authorizing family members to serve as surrogate decision makers, for example, typically include a limited list of family members in priority order.[2]

II. WHY THE FAMILY?

Several reasons support giving families privileged roles as surrogate medical decision makers. This section surveys these reasons, in light of three questions. First, why does the reason support a special role for the family as surrogate decision maker? Second, how strong is the reason? More precisely, does the reason provide just an argument for consulting the family? Or, is it strong enough to support a presumption in favor of the family? Is that presumption rebuttable or irrebuttable? Third, what is the nature of the family's role? Is the family involved as a source of information? As decision maker in terms of the patient's interests? As patient advocate? The survey of reasons begins with patient-centered concerns and then turns to family-centered reasons and the interests of society.

---

1. For a discussion of a functional approach to defining the family, see Martha Minow, *Redefining Families: Who's In and Who's Out?*, 62 U. COLO. L. REV. 269 (1991).

2. Utah's statute, for example, specifies the following ordered list of surrogate decision makers: holder of special power of attorney appointed by the patient, previously appointed legal guardian, spouse (if not legally separated), parents or surviving parent, child at least 18 years old (or a majority of reasonably available children at least 18 years old), nearest reasonably available living relative at least 18, and legal guardian appointed for the purpose of the health care decision at issue. UTAH CODE ANN. § 75-2-1105(2)(b) (Supp. 1992). An innovative New York proposal is to allow family members to agree upon a designated surrogate to replace the order which would otherwise apply. NEW YORK STATE TASK FORCE ON LIFE AND THE LAW, WHEN OTHERS MUST CHOOSE: DECIDING FOR PATIENTS WITHOUT CAPACITY app. A, at 252-53 (1992) [hereinafter NEW YORK STATE TASK FORCE].

A. The Patient

As the theory of decision making for incompetent patients has been developed, two standard approaches have emerged: an approach that aims to reproduce the patient's own choices and an approach that aims to further the patient's best interests. The attempt to reproduce the patient's own choice is seen as a recognition of the continued autonomy of a formerly competent patient and is preferred when the patient's choices can be known or reconstructed. The "best interests" approach is preferred for patients whose choices are unknown or for patients who have never been able to make competent choices.[3] Either of these approaches to decision making for incompetents requires accurate information, a need that may lead to the family.

1. Autonomy and Substituted Judgment

Suppose that the aim is to try to replicate what the incompetent patient would have chosen. Then, one important source of information is the patient's earlier directions about future eventualities. Only a small percentage of patients take advantage of formal legal mechanisms for directing health care in advance. However, with recent publicity and the Patient Self-Determination Act ("PSDA")[4] this percentage may increase. If there are reasons to doubt the accuracy of earlier directives, if the earlier record is nonspecific or vague, or if there is no earlier record, information will be needed that permits reconstruction of what the patient might have chosen.[5] For this reconstruction, it may be helpful to

---

3. *See, e.g., In re* Jobes, 529 A.2d 434, 452-61 (N.J. 1987) (Handler, J., concurring) (examining "substituted judgment" and "best interest" principles); ALLEN E. BUCHANAN & DAN W. BROCK, DECIDING FOR OTHERS: THE ETHICS OF SURROGATE DECISION MAKING 122-33 (1989) (discussing "best interest" standard). John Robertson has recently raised provocative questions about whether the choices of the formerly competent patient ought to be favored over the best interests of the now incompetent patient. John A. Robertson, *Second Thoughts on Living Wills*, HASTINGS CENTER REP., Nov.–Dec. 1991, at 6.

4. Patient Self-Determination Act, Pub. L. No. 101-508, § 4206, § 4751, 104 Stat. 1388, 1388-115, 1388-204 (1990) (codified at 42 U.S.C.A. §§ 1395cc, 1396a (West 1992)). For a discussion of difficulties with utilization of advance directives and implementation of the PSDA, see John La Puma et al., *Advance Directives on Admission: Clinical Implications and Analysis of the Patient Self-Determination Act of 1990*, 266 JAMA 402 (1991).

5. For a discussion of some of the ethical difficulties in allowing an earlier directive

have anecdotes of the patient working through actual or hypothetical medical decisions. More general information about the patient's values, methods of assessing information, or attitudes towards decision making could also be utilized in the reconstruction.[6]

In the litigated cases that attempt to reconstruct what the patient would have chosen,[7] families often report earlier discussions with the now-incompetent patient about health care decision scenarios. These discussions frequently are reported as reactions to the illnesses of other family members or friends. For example, when her husband was hospitalized with cancer, Mary O'Connor told her daughter that "she never wanted to be in a similar

---

to guide a patient's later care, see BUCHANAN & BROCK, *supra* note 3, at 152-89; Robertson, *supra* note 3, at 6-9. For a discussion of the possible vagueness of advance directives and a proposal for more precision, see Linda L. Emanuel & Ezekiel J. Emanuel, *The Medical Directive: A New Comprehensive Advance Care Document*, 261 JAMA 3288 (1989). For a discussion of physicians' and spouses' potential inaccuracy in predicting preferences of ill patients, see Allison B. Seckler et al., *Substituted Judgment: How Accurate are Proxy Predictions?*, 115 ANNALS INTERNAL MED. 92, 95-97 (1991); Richard F. Uhlmann et al., *Physicians' and Spouses' Predictions of Elderly Patients' Resuscitation Preferences*, 43 J. GERONTOLOGY M115 (1988). For a discussion of advance directive statutes and the easy revocation of these documents, see Leslie P. Francis, *The Evanescence of Living Wills*, 24 REAL PROP., PROB. & TR. J. 141 (1989). To take one typical example of how a state handles some of these difficulties, the Utah statute provides both that the desires of a competent declarant override an advance directive, UTAH CODE ANN. § 75-2-1108 (Supp. 1992), and that an advance directive can be revoked by oral statements of intent in the presence of a witness over 18 years of age who then signs a dated written instrument documenting the expression of intent. *Id.* § 75-2-1111(1)(c). As advance directives come into more general use, particularly if the PSDA encourages patients to complete advance directives under less than ideally thoughtful circumstances, it can be expected that there will be more doubts about their accuracy, and that family members will be a likely source of these doubts.

6. *See, e.g.*, Eric T. Juengst & Carol J. Weil, *Interpreting Proxy Directives: Clinical Decision-Making and the Durable Power of Attorney for Health Care*, *in* ADVANCE DIRECTIVES IN MEDICINE 21-37 (Chris Hackler et al. eds., 1989).

7. The legal term characterizing this effort is "substituted judgment." *See, e.g.*, *Jobes*, 529 A.2d at 456-57 (Handler, J., concurring) (discussing theory of substituted judgment). Sometimes courts are confused about this test, applying it to those who have never been competent. For example, in Superintendent of Belchertown State School v. Saikewicz, 370 N.E.2d 417 (Mass. 1977), the court applied the substituted judgment standard in deciding to withhold chemotherapy from a 67-year-old man profoundly retarded since birth. The *Saikewicz* decision has been widely criticized for confusing the purpose of substituted judgment. *See* BUCHANAN & BROCK, *supra* note 3, at 113-15. The pure idea of substituted judgment is to try to reconstruct what the patient would have chosen from what is known about how the patient made decisions. The surrogate decision maker, in exercising substituted judgment, tries to stand in for the incompetent by imaginatively reconstructing what the patient would have chosen from the information that is available about the patient. *Id.* at 117-22. To the extent that little information is known, this process becomes increasingly speculative and may be given less weight as a result. *Id.*

situation and that she would not want to go on living if she could not 'take care of herself and make her own decisions.'"[8] After her own heart attack, Mrs. O'Connor also told her daughter that "'she was very glad to be home, very glad to be out of the hospital, and hope[d] she would never have to be back in one again and would never want any sort of intervention[,] any sort of life support systems to maintain or prolong her life.'"[9]

Other typical examples of family reports involve discussions about the publicized but abstract fates of strangers. In *In re Swan*,[10] a case of a seventeen-year-old in a persistent vegetative state after an automobile accident, the patient's mother described a conversation with her son about a highly publicized termination-of-treatment decision. In discussing what it meant to be a "'vegetable,'" she explained that it meant requiring total care. Swan replied, "'If I can't be myself . . . no way . . . let me go to sleep.'"[11] In a more concrete but similar account, Swan's brother told of their visit to a comatose friend just eight days before his brother's accident. Reacting to the visit, Swan had remarked to his brother that "I don't ever want to get like that . . . . I would want somebody to let me leave—to go in peace.'"[12]

Somewhat vaguer reports are found in the *Jobes* case. Nancy Jobes is described as telling her first cousin that she wouldn't want "heroic measures" taken in the case of an accident and as telling her husband that "she would not want to be kept alive under Karen Quinlan's circumstances."[13] In still other cases, family members relate general perspectives on medical care—"she disliked going to doctors"[14]—or style of living—"Bertha Colyer was a very independent woman."[15] Family members may also report religious convictions and their guidance for decisions about care.[16]

---

8. *In re* Westchester County Medical Ctr., 531 N.E.2d 607, 611 (N.Y. 1988).
9. *Id.*
10. 569 A.2d 1202 (Me. 1990).
11. *Id.* at 1205.
12. *Id.*
13. *Jobes*, 529 A.2d at 442.
14. *In re* Welfare of Colyer, 660 P.2d 738, 748 (Wash. 1983).
15. *Id.*
16. *E.g.*, Brophy v. New England Sinai Hosp., Inc., 497 N.E.2d 626, 632 (Mass. 1986) (stating that surrogate decision maker "went through long and agonizing research, reflection, and prayer," and discussed decision with clergy).

Some courts are quite willing to rely on these familial stories. For example, in authorizing the withdrawal of treatment from Bertha Colyer, the Washington Supreme Court commented: "Given the unanimity of the opinions expressed by Bertha's closest kin, together with the absence of any evidence of any ill motives, we are satisfied that Bertha's guardian was exercising his best judgment as to Bertha's personal choice when he requested the removal of the life support system."[17]

Other courts, however, express doubt about the specificity or reliability of these familial stories as a basis for understanding what the patient would have chosen. The New York court, for example, pointed out that Mary O'Connor had not explicitly discussed medically assisted nutrition and hydration with her daughter; nor had she expressed attitudes towards pain or comfort care.[18] In the *Jobes* case, the New Jersey court refused to rely on the family's evidence as a basis for a reconstruction of what Nancy would have chosen: "All of the statements about life-support that were attributed to Mrs. Jobes were remote, general, spontaneous, and made in casual circumstances. Indeed, they closely track the examples of evidence that we have explicitly characterized as unreliable."[19]

Courts that insist on clear and convincing evidence of the patient's wishes before a termination of treatment decision can be effectuated, such as New York, may be especially skeptical about the status of these familial reports.[20] But skepticism is not limited to courts which insist on a high evidentiary standard; a California Court of Appeals, for example, has pointed out the unreliability of the patient's informal statements, made perhaps years earlier, when constitutional rights are involved: "It would be a dangerously unpredictable precedent."[21]

Thus, when the effort is to replicate what the patient would have chosen and thereby recognize the patient's autonomy, the reason for family involvement is the likelihood that the family will

---

17. *Colyer*, 660 P.2d at 748.

18. *Westchester County Medical Ctr.*, 531 N.E.2d at 611.

19. *Jobes*, 529 A.2d at 443.

20. *E.g.*, Cruzan v. Harmon, 760 S.W.2d 408, 424 (Mo. 1988) (en banc) (stating that similar to *Jobes*, statements used to determine Nancy Cruzan's intent are unreliable) (citing *Jobes*, 529 A.2d at 443).

21. Drabick v. Drabick, 245 Cal. Rptr. 840, 856 (Ct. App. 1988).

have good information about the patient's choices, preferences, or values. In these cases, the reason for family involvement is only as strong as the likely evidentiary value of the family's knowledge. When family members are the only good sources of this kind of information, they may appear uniquely positioned to help in decision making. But this unique positioning is contingent on the accuracy and exclusiveness of the family's information. Thus, the family's unique position provides at best a good reason for consulting the family but not an irrebuttable presumption in favor of family involvement. On the other hand, if the patient has exercised a durable power of attorney to select a family member (or someone else) as decision maker in the case of incompetence, there would be a much stronger reason for recognizing the authority of the surrogate, rebuttable only by a showing that the appointment of the surrogate itself was flawed.[22]

2. The Patient's Best Interests

Information about the patient's interests is required for application of the best interests standard. Here, too, the family may have important information. Court decisions applying the "best interests" standard have done little to provide a general theoretical account of what they mean by "interests." Perhaps the fullest judicial account was given by the Arizona Supreme Court: interests involve "such objective criteria as relief from suffering, preservation or restoration of functioning, and quality and extent of sustained life."[23] The more theoretical account relied on by the Arizona court was provided by the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research: "An accurate assessment [of interests] will encompass consideration of the satisfaction of present desires, the opportunities for future satisfactions, and the possibility of developing or regaining the capacity for self-determination."[24] The

22. For a recognition of the importance of special powers of attorney for health care, see Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2857-58 (1990) (O'Connor, J., concurring).

23. Rasmussen v. Fleming, 741 P.2d 674, 689 (Ariz. 1987) (en banc) (footnote omitted).

24. PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, DECIDING TO FOREGO LIFE-SUSTAINING TREAT-

ability to experience pleasure and the receipt of health care with the potential to restore function are typically regarded as in a patient's interests, and the experience of pain or uncomfortably intrusive health care are typically regarded as not in a patient's interests.[25]

The family may be able to provide information about the patient's wants and, in some cases, about what means will help to satisfy them. In reported decisions, families often indicate the extent to which continued care is painful or disturbing to the patient or, on the other hand, is well tolerated. For example, from the age of five, Barbara Grant suffered from Batten's disease, a genetic disease causing progressive neurological deterioration. Her mother, seeking to have life support withheld during the final stages of the illness, described her daughter's dislike for taking medicine, using a cane, having suction tubes used on her, and her dislike for the medical staff.[26] In another reported decision, John Storar was fifty-two years old, with a mental age of about eighteen months.[27] He suffered from bladder cancer and was expected to live four to six months with regular blood transfusions. His seventy-seven-year-old mother, who lived nearby and visited him daily, observed that the transfusions disturbed him and requested to have them discontinued.

Despite the likelihood that family members will have information that is crucial to assessing what is in the patient's interests, application of the best interests standard also may require information that is not particularly likely to be within the scope of the family's special knowledge. For example, the standard requires an objective assessment of the benefits and risks for the patient of continued care, an assessment that may require quite technical medical understanding.[28] In applying the standard, courts thus may give weight to medical testimony about the likely results of treatment. To the extent that an assessment of interests requires factual judgments that are beyond the family's particular expertise, courts may be less likely to turn to the family and more

---

MENT RESEARCH 135 (1983).

25. *See Rasmussen*, 741 P.2d at 689.

26. *In re* Guardianship of Grant, 747 P.2d 445, 448 (Wash. 1987) (en banc).

27. *In re* Storar, 420 N.E.2d 64, 68 (N.Y. 1981).

28. *See, e.g., Jobes*, 529 A.2d at 457 (Handler, J., concurring) (decision maker must consider needs, risks, and benefits to affected person).

likely to turn to expert sources, particularly physicians, for relevant information.[29]

In addition to their knowledge, family members may also be better motivated than others to be sure that standards for decision making—either substituted judgment or the patient's best interests—are applied carefully and accurately. Family members may be motivated to seek out information about the patient's expressed preferences. Similarly, they may be more motivated than others to pursue the information needed to decide what is in the patient's interests—for example, information about various sources of financing for care or about alternative facilities for treatment.

The *Jobes* decision is an excellent example of a court's reliance on the family's care and concern for the patient:

> Family members are best qualified to make substituted judgments for incompetent patients not only because of their peculiar grasp of the patient's approach to life, but also because of their special bonds with him or her. Our common human experience informs us that family members are generally most concerned with the welfare of a patient. It is they who provide for the patient's comfort, care, and best interests and they who treat the patient as a person, rather than a symbol of a cause.[30]

As the *Jobes* court notes, familial caring may be directed to the needs of the individual patient, rather than more abstractly towards a value such as preserving life or providing good medical care. Caring may also motivate families to be persistent advocates for the patient's interests, doggedly insisting that the patient receive attention, comfort, respect, and any care that might prove beneficial.

In addition to their knowledge and motivation, families are likely to be important to the success of various forms of care for incompetent patients. Families are the most likely caregivers for

---

29. *See, e.g., id.,* at 460 (Handler, J., concurring) (suggesting that in ambiguous cases, decision maker should consult doctors, government and institutional representatives, and people with religious or ethical training). For a defense of this approach, see Michele Yuen, Comment, *Letting Daddy Die: Adopting New Standards for Surrogate Decisionmaking,* 39 U.C.L.A. L. REV. 581, 608-631 (1992).

30. *Jobes,* 529 A.2d at 445 (citations omitted).

the debilitated. Their help is also likely to be needed in implementing various forms of therapy, particularly as patients are discharged more quickly from care facilities. For example, the participation of family members may be important to rehabilitation programs for stroke victims or educational programs for disabled children. Thus, if families become invested in the decisions about care, these decisions are more likely to work out well for patients. Each of these links to the family, however—knowledge of interests, care and concern, and involvement in therapy—are contingent and may well not apply in specific family circumstances.

B. The Patient's Future

To this point, the discussion of patient-centered reasons has assumed a patient with settled choices, preferences, or interests—someone who has become a developed person, although perhaps only to an extent limited by medical events. But some patients, especially young children, do not have a full range of established choices or even readily predictable interests. There is not as yet a template of choices or interests of the patient on which to base decisions. The template remains to be developed, if possible. Health care decisions, like other decisions that affect the fates of young children, will shape preferences and interests and thus will help create the person the child becomes. Decisions for children therefore may entail weighing options for very different kinds of persons and lives.

This contrast between adults and very young children is, to be sure, a matter of degree rather than absolute. As children mature, their interests become clearer and more settled. Most young children do have interests related to occurrent preferences, as well as to how they will develop, for example, interests in being fed or free from pain. On the other side, the interests of adults are open to change in the future and will to some extent be shaped by the medical choices that are made. Therapeutic options that result in loss of a limb, damage to sight or hearing, or infertility, may open—or close—very different futures for patients, and their choices and interests may shift in response. Nevertheless, for adult patients these choices take place against a template of already-shaped preferences about physical activity, music, or bearing

children. For young children, the individualized template is far less clear.

This contrast between the relative open-endedness of the preferences of children and the relative development of the preferences of adults has important implications for the role of the family in health care decision making. The approach to adult decision making, as it has been developed in contemporary bioethics and as it was discussed in the preceding section, is grounded in several important assumptions of liberal theory. The starting place for analysis is the individual (in this case, the patient). A very important moral task (if not the most important moral task) is to respect the patient's autonomy as far as possible.[31] Autonomy is respected either by letting the patient choose, or by relying upon an already-developed template of values, preferences, interests, and choices. When autonomy is not a possibility, the patient is assumed at least to have individualized interests that are to be the basis of decision making.

For children, in contrast, the background template is yet to be constructed. A theory of health care decision making must include an account of how the background template is to be filled in, including a view about the roles for parents or other family members. There are, of course, many different views about the roles of parents in shaping the preferences and interests of their children. One basic division is whether the claims of parents or the claims of children are the most fundamental to a theory of parent-child relationships.[32] Another is the meaning and value of autonomy for the developing child.[33] For those who would regard autonomy as an important value, still another issue is the role of parents in fostering autonomy. With respect to parent-child relations, perhaps the prevailing liberal views are that the claims of the individual child are foundational and that children should

---

31. The canonical liberal text in contemporary bioethics likely is THOMAS L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS (3d ed. 1989).

32. Perhaps most fundamentally, these views differ on whether the parents' claims or the child's claims are theoretically prior. *See* JEFFREY BLUSTEIN, PARENTS & CHILDREN: THE ETHICS OF THE FAMILY 104-14 (1982). It is worth noting that these discussions tend to consider the role of parents, rather than family members more generally.

33. *See, e.g.*, WHOSE CHILD? CHILDREN'S RIGHTS, PARENTAL AUTHORITY, AND STATE POWER (William Aiken & Hugh LaFollette eds., 1980) [hereinafter WHOSE CHILD?] (presenting essays on relationship between child, parent, and state).

be raised to become autonomous adults. These liberal assumptions are severely criticized, particularly by those who doubt the individualism and autonomy on which they rest.[34] Nevertheless, these assumptions are a useful starting place because they represent the corresponding theory about decision making for children that the prevailing liberal theory in bioethics represents about decision making for adults. So the aim in what follows is to draw out some of the implications of this liberal theory for the role of parents in making health care decisions for their children. The aim, in short, is to outline a liberal account of the role of parents in making health care decisions for their children, not to defend liberal theory more generally.

The child-centered views of parental obligations that have predominated in recent liberal political theory base parents' obligations towards their children on a vision of what their children should become: autonomous adults, capable of choosing the kinds of lives they want to lead. Joel Feinberg describes this theory as the "right to an open future."[35] By this, he means that children have rights to have certain "key options" continue to be available for them, such as choices of careers that fit their talents and dispositions, until they are able to make choices among them as adults. Parents, in turn, are obligated to try to create the conditions that help children realize their rights to open futures. Thus, according to the liberal view, parents are obligated to try to provide their children with the sustenance, support, and education needed for growth to autonomous adulthood.

For its critics, perhaps the most controversial aspects of this liberal view are its incompatibility with parental values and traditional forms of social life. But the theory is subject to criticism even on its own terms: the creation of the conditions for eventual self-determination may not be a value-free enterprise in itself; it may, by creating the conditions for one sort of life, effectively preclude the enjoyment of others. We might say that choices about children's lives are *neutral* towards their futures, to the extent that they leave options open for later determination by the child

---

34. *See* MICHAEL J. SANDEL, LIBERALISM AND THE LIMITS OF JUSTICE *passim* (1982).

35. Joel Feinberg, *The Child's Right to an Open Future*, *in* WHOSE CHILD?, *supra* note 33, at 126; *see also* BLUSTEIN, *supra* note 32, at 199 (asserting that parents have duty "to expose their children to the psychological conditions that facilitate the development of their capacities for self-determination, or autonomy").

as an adult. The extent to which creation of the conditions for self-determination is neutral in this sense is highly controversial. For instance, some liberals tend to believe that this neutrality does and should have a relatively wide range.[36] Writers more critical of the liberal tradition are doubtful.[37]

Doubts about both the possibility and the desirability of neutrality are clearly illustrated by controversies about education, and these controversies are instructive for decision making about health care.[38] There are highly charged questions about the extent to which education that helps create the conditions for self-determination is consistent with parental influence about values. Are parents neutral if they encourage their children to love music, sports, or God? Can any of these attitudes be "revoked," and, if not, does liberal neutrality recommend confining education to bloodless values that instill no passions that the child cannot later overthrow? Would it, for example, violate a child's right to an open future to bring the child up in a very structured religion, which inculcates significant feelings of guilt for abandonment of the faith?[39] To limit the child's education in order to avoid contact with the "modern" world, as the Amish do?[40] To encourage a child to pursue competitive gymnastics, a career as a concert violinist, or some other calling that may require an overwhelming commitment from an early age? To conclude that parents may not choose any of these pathways for a child because the choice is not neutral among futures imposes drastic limits on what parents may do for their children: they cannot create deep concerns, loves, or commitments that are inconsistent with an open future. If the

---

36. Feinberg, *supra* note 35, at 124-51.

37. Sharon Bishop, *Children, Autonomy, and the Right to Self-Determination, in* WHOSE CHILD?, *supra* note 33, at 154-76.

38. This is a problematic analogy for those who believe that one set of values governs education—for example, the duty to instruct children in the ways of God—but a different set of values governs health care decision making, such as the preservation of life at all costs. With views such as these, the parents are seen as agents of the values in question. The perspectives which include these values may, however, also see the parents as preferred agents for value transmission. *See, e.g.*, *The Country's Future Is in Your House,* WASH. POST, Aug. 20, 1992, at A34 (excerpts from Barbara Bush's speech at Republican National Convention indicating that parents are most important vehicle for transmitting values to children).

39. *Cf.* JAMES JOYCE, A PORTRAIT OF THE ARTIST AS A YOUNG MAN (Penguin Books 1983) (1916) (illustrating these religious conflicts from Catholic perspective).

40. *Cf.* Wisconsin v. Yoder, 406 U.S. 205, 210-12, 236 (1972) (upholding right of Amish parents to remove their children from school after eighth grade).

right to an open future is understood to require extensive neutrality among courses that cannot later be rejected, the theory seems implausible as a complete account of the role of parents in their children's lives.

In the debates about education, however, the right to an open future is not understood in this pallidly neutral way. Suppose, then, that we try a somewhat fuller characterization of which future options should be kept open for a child. Three quite tentative suggestions can be drawn from remarks made by liberal theorists such as Feinberg. First, actions taken during childhood should not be ones that would violate the rights of the later adult. If adults generally enjoy political rights such as the right to vote, rights of personal choice such as the right to marry or to choose whether to bear children, or liberty rights such as freedom of speech, parents should not take steps during childhood that would violate these later rights. Thus, for example, parents should not be able to renounce a child's rights of citizenship.

Second, actions taken during childhood should attempt to uncover and foster a reasonable measure of the child's abilities and talents.[41] Thus, parents should not impose educational or work regimes that offer the child little opportunity to discover talents or that ruthlessly attempt to discourage significant abilities.[42] On the other hand, children are multifaceted and not all talents can be developed simultaneously; part of what is involved in the openness of the child's future is the need for guidance and selection among various possibilities for development. Parents may, along with their children as they become older, select among reasonable sets of these possibilities.

---

41. Feinberg, for example, links the child's "open future" to the development of talents:

> [T]he parents who raise their child in such a way as to promote his self-fulfillment most effectively will at every stage try to strengthen the basic tendencies of the child as manifested at that stage. They will give him opportunities to develop his strongest talents, for instance, after having enjoyed opportunities to discover by various experiments just what those talents are.

Feinberg, *supra* note 35, at 150.

42. On these grounds, for example, Feinberg is much more doubtful about the permissibility of the Amish decision to forego public education after the eighth grade, Wisconsin v. Yoder, 406 U.S. 205 (1972), than about the permissibility of Jehovah's Witnesses using their children to distribute pamphlets on street corners, Prince v. Massachusetts, 321 U.S. 158 (1944). *See Feinberg, supra* note 35, at 129-38.

Third, actions taken during childhood should not preclude the later adult from enjoying major life satisfactions, such as deep human relationships.[43] This third suggestion is difficult to formulate and to balance against the second. There are certainly tensions between them. Sharon Bishop, for example, argues that education of a female child that is aimed at developing talents is crucial for autonomy even though it may conflict with traditional female roles.[44] These tensions are at the heart of the dialogue about an education that develops understanding and intellectual talents, at the likely expense of deeply held communitarian values.

These three suggestions about how to understand the openness of the child's future in the context of education might be applied to medical decision making as follows. First, decisions that would violate the rights of the later adult would be prohibited. For example, if the later adult has rights to procreate, sterilization of the child would violate these rights. If the later adult has rights to nurture her own children, compelled abortion on an adolescent patient would violate those rights. Second, in making medical decisions, parents should try to the extent possible to take into account the development of reasonable groupings of children's abilities and talents. They should not, for example, deny care or delay care when to do so carries significant risks of mortality or morbidity. Nor should they choose therapies that unreasonably risk compromising cognitive capacities or major functional possibilities. On the other hand, parents may legitimately weigh risks to one sort of capacity against risks to another—for example, choices between chemotherapeutic modalities that weigh an increase in the chance of limb salvage against sterility or hearing loss. Third, parents should not make decisions, if at all possible, that are likely to preclude important human satisfactions. For example, in medical decision making they should take into account the preservation of communicative and perceptual capacities.[45]

---

43. BLUSTEIN, *supra* note 32, at 199, argues that parents have twin duties towards their children: raising them to be capable of self-determination and promoting their self-fulfillment.

44. Bishop, *supra* note 37, at 154-76.

45. For a criticism of this suggestion and defense of the view that parents should not be required to foster the development of life prospects that they find unacceptable, see William Ruddick, *Parents and Life Prospects*, *in* HAVING CHILDREN: PHILOSOPHICAL AND LEGAL REFLECTIONS ON PARENTHOOD 124-37 (Onora O'Neill & William Ruddick eds., 1979). Ruddick gives the example of dwarf parents who very much want to raise a dwarf

These suggestions outline the latitude of parental discretion in health care decision making. They would not, for example, require parents to prolong life if doing so would not be helpful in allowing the child to develop talents or experience important life satisfactions. Thus, to this extent, parents would be permitted to make quality of life judgments for their children, including judgments to forego treatment when the prognosis is dim for the child's ability to develop talents or to experience well-being at even a minimal level. Parents would also be permitted to choose among alternate possible lives consistent with these guidelines and to weigh therapeutic alternatives in light of the different risks to functional capacities that they present. For example, parents could choose therapeutic options that risk death in exchange for significant improvements in functional status, weigh options that trade different kinds of compromises in functional status against each other (for example, surgery as against chemotherapy for certain types of cancers or choice of gender for a child born with ambiguous gender characteristics), or weigh significant compromises to well-being against other possible advantages of care.

On this understanding of the implication of liberal theory for parent-child relationships, parents thus have significant latitude in shaping the futures of their children. This latitude embodies choices among forms of health care that may shape quite different futures, including even the possibility of death. Furthermore, the role of parents is not simply informational or contingent; it is an integral part of the decision, as long as the decision is exercised in accord with the constraints suggested above, unless some circumstance disqualifies the parents as parents altogether.

C. Interests of the Family

Within patient-centered models of medical decision making, the interests of the family are relevant only insofar as they make a difference to the patient's choices or interests. As other Articles

---

child and refuse medical care that would cause the child to be of normal stature. He concludes that the parents are not obligated to provide the care, despite the predictability that being of normal size would open important life prospects for the child: "The parents . . . are violating the child's right to a normal life—if there is such a right. I see no basis for such a right, nor do I think this case requires us to look for one. If there is such a right, we may set against it the dwarves' right to be parents." *Id.* at 133-34.

in this symposium indicate, entirely patient-centered models may be increasingly regarded as myopic, but this is not the point here. The models of medical decision making developed in this Article are based largely on the interests of the patient, but it is important to survey briefly the separate interests of family members because they are often among the most important factors in actual decision making.

First, it is obvious that family members' medical decisions may have remarkably intense and long-lasting effects on family emotions, relations, time, opportunities, and finances. The care of a debilitated relative can be a daily intrusion on marital relationships, a drastic limit on career choices, or a perceived barrier to having (other) children. Although adoption of or reliance on alternative sources of care are readily proposed as alternatives, neither option is easily taken without emotional pain. Various social arrangements in the United States particularly complicate the burdens. For example, Medicare pays only very limited home health benefits,[46] and there are economic difficulties in qualifying for Social Security disability benefits.[47] If consequences for the family are considered relevant to health care decision making, it certainly seems that there will be cases in which the costs to the family will be so great as to outweigh any contemplated benefit of continued care for the afflicted family member.

Second, views about the moral importance of relationships may support an enhanced role of the family in medical decision making for incompetent family members. For example, if the potential for a parent-child relation is regarded as a very important object of care, then the family ought to have a role in deciding about care as it affects this relationship. For those who place the relationship rather than the individual at the center, the family is integral to the process of health care decision making.

III. THREE DIFFERENT ROLES FOR THE FAMILY

The patient's choices, interests, and open future thus are important patient-centered characteristics that support roles for the family in making medical decisions for incompetent members.

---

46. *See* 42 C.F.R. § 410.10 (1991).
47. *See* 20 C.F.R. § 416.1100–.1182 (1992).

When the aim of decision making is to attempt to capture or recapture the patient's choices, family members frequently will be important sources of information. They may know of actual choices or be able to report discussions about desires in hypothetical situations. They may be able to supply important background information about the patient's values or styles of decision making and the extent to which these have been long-lasting and consistent. On the other hand, the family may have been at worst estranged and at best simply uncommunicative about medical decision making. Thus, when there is some possibility of reconstructing the choices of a formerly competent family member, the primary role for the family is to serve as a critical source of information about the patient.

When the aim of medical decision making is realization of the patient's best interests, either because the patient's choices cannot be reconstructed or because the patient has never been competent, the family may also be an important source of information. The family may have knowledge about how intensely a patient felt various satisfactions or how much discomfort the patient apparently suffered from illness or therapeutic intervention. On the other hand, a critical part of the assessment of a patient's interests is understanding of the medical prognosis and alternatives for care, and in this respect the family is not in a privileged position. Because of its love and concern, however, the family may be the most likely agent to press for consideration—and, importantly, for reassessment and reconsideration—of the patient's interests. Thus, when the best interests standard is the most appropriate one for medical decision making, the family's principal role is as advocate for the patient's interests.

Finally, when the patient's interests have yet to be developed significantly, medical decisions may shape that development in important ways. By analogy to the role of the parent in education, parents may make some of these seminal health care decisions for their children, subject to three important guidelines. First, parents should not act in ways that violate the rights of the adult the child will become. Second, parents should try to ensure the continued availability of a life in accord with a reasonable range of the child's talents and abilities. Third, parents should try to preserve important capacities for their children to experience satisfactions. In making medical decisions for their children, therefore, the

parents' role extends to shaping, as well as advocating, the children's interests. This is particularly true for infants and younger children: as children become increasingly capable of shaping their own lives autonomously, the roles of family members increasingly shift towards their roles for adult patients.

These roles for the family—informant, advocate, and shaper—are supported within the predominant liberal paradigm of bioethics. If liberal assumptions such as the priority of the individual or the importance of autonomy are discarded, then there may be different and even stronger roles for the family.

IV. CURRENT LEGAL APPROACHES: THE CONTRAST BETWEEN ADULTS AND YOUNG CHILDREN

The last ten years have seen remarkable statutory and case-law development with respect to medical decision making for incompetent patients. Insofar as trends are discernable, the legal developments seem to more clearly allow discretion for familial decision making when the patient is an adult than when the patient is a young child. This divergence seems contrary to the suggestions drawn from the liberal models about the ethics of health care decision making.

A. Adults

With respect to adults, case law has tended to allow family members relatively wide latitude in reconstructing patients' choices. A good example is the Washington Supreme Court's decision in *Colyer*, which relied on the family's reconstruction of Bertha Colyer's values of independence and distrust of medical care.[48] Even the *Cruzan* case was ultimately resolved in accord with the family's wishes.[49] Despite the state's insistence that there be clear and convincing evidence of the patient's wishes, the trial court eventually found that the family had brought forward sufficient evidence to show that Nancy would not have wanted her

48. *In re* Colyer, 660 P.2d 738, 748 (Wash. 1983) (en banc).

49. Diane E. Hoffman, *Introduction: The Right to Die After Cruzan*, 2 MD. J. CONTEMP. LEGAL ISSUES 93, 96 (1991).

life prolonged in a persistent vegetative state by means of medically assisted nutrition and hydration.[50]

In addition to the court decisions, some states now also have statutes that authorize family members to consent to the withdrawal of life sustaining treatment, most frequently when the patient is terminally ill.[51] These statutes typically list family members in an automatic order of priority, after the patient's own choice of a surrogate or a court appointed guardian. New York's proposal to let family members themselves select the surrogate is innovative.[52] Although these provisions are often part of statutes establishing living wills or special powers of attorney for health care, they typically enumerate family members by degree of relationship rather than by knowledge or known intimacy to the patient. One justification for the preset statutory priorities is that they are highly likely to mirror the justifications for reliance on the family, especially knowledge and caring. But there is no automatic requirement to ascertain whether this is so in any particular case. Instead, if there is significant disagreement with the statutory ordering of family members, the statutory alternative is to seek appointment of a court-appointed guardian who then takes priority. The most likely entity to pursue guardianship

---

50. *Id.*

51. States with family consent statutes include: Arkansas (ARK. CODE ANN. § 20-17-214 (Michie Supp. 1991)); Connecticut (CONN. GEN. STAT. § 19a-571 (1991)); the District of Columbia (D.C. CODE ANN. § 21-2210 (1989)); Florida (FLA. STAT. ANN. § 765.07 (West 1986)); Hawaii (HAW. REV. STAT. § 327D-21 (1991)); Idaho (IDAHO CODE § 39-4303 (1985) (medical consent statute)); Illinois (ILL. ANN. STAT. ch. 110½, para. 851-25 (Smith-Hurd Supp. 1992)); Indiana (IND. CODE ANN. § 16-8-12-4 (Burns 1990), construed in *In re* Lawrence, 579 N.E.2d 32, 41-43 (Ind. 1991)); Iowa (IOWA CODE ANN. § 144A.7 (West 1989)); Louisiana (LA. REV. STAT. ANN. § 40:1299.58.5 (West 1992)); Maine (ME. REV. STAT. ANN. tit. 18-A § 5-707(b) (Supp. 1991)); Maryland (MD. CODE ANN., HEALTH-GEN. § 20-107(d) (Supp. 1991) (medical consent statute)); Mississippi (MISS. CODE ANN. § 41-41-3 (Supp. 1991)) (medical consent statute); Montana (MONT. CODE ANN. § 50-9-106 (1991)); Nevada (NEV. REV. STAT. ANN. § 449.626 (Michie 1991)); New Mexico (N.M. STAT ANN. § 24-7-8.1 (Michie 1986)); New York (N.Y. PUB. HEALTH LAW § 2965 (McKinney Supp. 1992) (limited to DNR orders)); North Carolina (N.C. GEN. STAT. § 90-322 (1990)); Oregon (OR. REV. STAT. § 127.635 (1991)); South Dakota (S.D. CODIFIED LAWS ANN. § 34-12C-3 (Supp. 1992) (medical consent statute)); Texas (TEX. HEALTH & SAFETY CODE ANN. § 672.009 (West 1992)); Utah (UTAH CODE ANN. § 75-2-1107 (Supp. 1992)); Virginia (VA. CODE ANN. § 54.1-2986 (Michie Supp. 1992)); and Washington (WASH. REV. CODE ANN. § 7.70.065 (West 1992)) (medical consent statute).

52. NEW YORK STATE TASK FORCE, *supra* note 2, app. A, at 253.

proceedings is the treating health care institution, motivated by fears of liability when family members disagree.[53]

B. Children

The legal situation for children is less developed and more diverse. Some of the earlier cases appeared to be moving towards the view that parents may not refuse life-saving treatment but have latitude to make decisions when there is some dispute about what is in their children's interests. Other cases, however, gave parents more latitude, particularly several highly controversial instances in which corrective surgery was withheld from infants with Down's syndrome. In the early 1980s, the Baby Doe regulations took a strong stand in favor of aggressive treatment of nearly all newborns.[54] Perhaps because of the regulations, case-law development was notably limited. Several states passed statutes codifying the Baby Doe regulations;[55] in most others, the legal status of parental decisions to withhold or withdraw care remains unclear.

1. Before Baby Doe

By the late 1970s, case law had appeared with respect to nontreatment decisions for infants and children. For the most part, the decisions involved parents who had refused standard medical recommendations on religious grounds.[56] When there was medical agreement that the recommended care had the clear potential to avoid mortality or morbidity, courts generally would mandate it.[57] Several cases left decisions within the parents' discretion

53. ALAN MEISEL, THE RIGHT TO DIE § 6.1–.27 (1989 & Supp. 1991)).

54. *See infra* notes 79-86 and accompanying text.

55. *See infra* notes 91-101 and accompanying text.

56. *See* MEISEL, *supra* note 53, § 13.5–.6.

57. *E.g.*, Morrison v. State, 252 S.W.2d 97, 103 (Mo. Ct. App. 1952) (holding that state has authority to mandate blood transfusion to preserve child's life); State v. Perricone, 181 A.2d 751, 756-57 (N.J. 1962) (affirming order of blood transfusion for infant over Jehovah's Witness parents' objection); Sampson v. Taylor (*In re* Sampson), 278 N.E.2d 918, 919 (N.Y. 1972) (holding religious objection by parent to blood transfusion was not bar to court order in neglect proceeding when transfusion was necessary to success of required surgery); Application of Brooklyn Hospital, 258 N.Y.S.2d 621, 623 (Sup. Ct. 1965) (granting hospital administrator authority to consent to child's blood transfusion when child was seriously endangered and parents objected to

when the court perceived significant dispute about the possible efficacy of the care or likely risks to the child.[58] Although relatively sparse, these cases seem to track the liberal view sketched above:[59] parents may choose among reasonable courses of action in shaping their children's futures, but they may not take avoidable risks of cutting off significant capacities or even life itself.

The relatively undeveloped state of the law was highlighted by two much discussed cases in 1979. In Massachusetts, the Chad Green case involved a parental decision to discontinue chemotherapy in favor of giving laetrile to their young son with leukemia.[60] Chad Green was twenty months old when his disease was first diagnosed; the treating physicians recommended chemotherapy with an apparent prognosis of a better-than-fifty-percent likelihood of five-year survival. The risks of chemotherapy, as described by the physicians, were relatively benign, easily managed side effects such as constipation. The parents, however, were very upset by the way the chemotherapy made their son feel and discontinued it.

The court ordered the parents to provide the care,[61] but the reasoning in the case was less than fully clear. In the first hearing, in which the state sought to compel continued therapy, the court balanced three factors: the natural rights of parents, the best interests of the child, and the interests of the state.[62] The factors were presented as a list, without a theoretical account of which should predominate or why. The court's decision rested on the conclusion that each of the factors pointed in the same direction. The rights of the parents, in the court's view, were to be treated as a trust which did not extend to the right to risk the life

---

transfusion on religious grounds); *In re* Clark, 185 N.E.2d 128, 132 (Ohio C.P. 1962) (stating that when child's right to live and parents' religious beliefs collide, former is paramount).

58. *E.g.*, *In re* Seiferth, 127 N.E.2d 820, 824 (N.Y. 1955) (allowing parents to decide about child's cleft palate surgery); *In re* Hudson, 126 P.2d 765 (Wash. 1942) (discussing amputation of grossly deformed arm to which parents objected because of risks of surgery). It may be significant that these are older cases, and the courts were less inclined to view medical care itself in a favorable light.

59. *See supra* notes 31-45 and accompanying text.

60. Custody of a Minor, 379 N.E.2d 1053 (Mass. 1978).

61. *Id.* at 1062.

62. *Id.* at 1061-67.

or health of the child.[63] The best interests of the child would clearly be served by therapy: the chemotherapy had minimal side effects (so the court found) and offered a good chance of cure; moreover, there were no other available alternatives.[64] And the state's interest was to protect the child.[65] In a later hearing, the court used the same analysis to order the parents to discontinue therapy with laetrile and vitamins in addition to ordering the chemotherapy.[66] This second conclusion illustrates the difficulty with the court's approach because the evidence about the risks of the supplemental therapy was quite thin, and the parents' hopes were simply set aside.

On the other hand, balancing the parents' rights against the child's interests can also lead to problematic latitude for parents. In a New York case parallel to the Chad Green fact situation, the court permitted the parents of a seven-year-old with Hodgkin's disease to refuse chemotherapy in favor of laetrile on the recommendation of a licensed New York physician who specialized in nutrition.[67]

The difficulty with listing the parents' rights along side the child's interests is starkly illustrated by a second decision in 1979, the California case of Phillip Becker. Born with Down's syndrome and a heart defect, Phillip Becker, at twelve years old, faced the prospects of increasing shortness of breath and ultimate lung failure by early adulthood if he did not have surgery to repair the heart defect.[68] Although the surgery had been recommended for several years, Phillip's parents refused their consent. The State petitioned for Phillip to be declared a dependent child of the court for the purpose of surgical consent. Despite Phillip's natural parents' failure to maintain extensive contact with him,[69] the court rejected the State's petition and allowed the parents to

---

63. *Id.* at 1063.

64. *Id.* at 1065.

65. *Id.* at 1066-67.

66. Custody of a Minor, 393 N.E.2d 836, 845-46 (Mass. 1979). The court concluded that the laetrile and vitamin therapy put the child at risk of low-grade cyanide poisoning and brain damage. *Id.* at 845. By the time the court actually heard the case, the parents had fled with the child. *Id.* at 838 n.1.

67. Saratoga County Dep't of Social Servs. v. Hofbauer (*In re* Hofbauer), 393 N.E.2d 1009, 1012-15 (N.Y. 1979).

68. Bothman v. Warren (*In re* Phillip B.), 156 Cal. Rptr. 48, 50 (Ct. App. 1979).

69. Herbert H. v. Warren B. (Guardianship of Phillip B.), 188 Cal. Rptr. 781, 786-87 (Ct. App. 1983).

refuse the surgery.[70] The court began its analysis with the autonomy rights of the parents:

> Inherent in the preference for parental autonomy is a commitment to diverse lifestyles, including the right of parents to raise their children as they think best. Legal judgments regarding the value of childrearing patterns should be kept to a minimum so long as the child is afforded the best available opportunity to fulfill his potential in society.[71]

At the same time, according to the court, the state has an interest in protecting children, and may thus interfere in family matters "to safeguard the child's health, educational development and emotional well-being."[72] The trial court denied the petition because the evidence did not show clearly and convincingly that the surgery was necessary to safeguard Phillip's health.[73] In affirming the trial court's decision, the appellate court agreed that the clear and convincing standard was proper.[74] Testimony at the trial court included evidence that the surgery was somewhat more risky for Down's syndrome patients than for other children (for whom it had a five to ten percent mortality risk), that there was the possibility of complications requiring a pacemaker, and that Phillip already might have suffered some lung damage from his heart condition. The appellate court characterized the trial court as balancing the benefits and risks of the surgery for Phillip,[75] a characterization that has been criticized as slanted towards the parents' conclusions.[76]

Intermittently throughout the 1970s, reports appeared about parental decisions to withhold care from ill newborns. A study of medical practice indicated that a number of infants with Down's syndrome or neural-tube defects died after nontreatment deci-

---

70. *Phillip B.*, 156 Cal. Rptr. at 52.

71. *Id.* at 51.

72. *Id.*

73. *Id.* at 52. The state, however, had urged the lower "preponderance of the evidence" standard. *Id.*

74. *Id.*

75. *Id.*

76. *See generally* Kathleen M. Heydon, Note, *Guardianship of Phillip B.: Nonparents' Right to Custody in California*, 18 LOY. L.A. L. REV. 779 (1985) (discussing history of Phillip Becker and analyzing court decisions).

sions.[77] Several highly publicized parental decisions to refuse care—especially the "Baby Doe" cases—brought the issue into the political forum. Several factors may have combined to explain the aggressiveness of the Baby Doe regulations: the growth of the right to life movement, the Reagan presidency, the uncertainty of case law, and changes in attitudes and understanding about Down's syndrome patients.[78]

2. The "Baby Doe" Regulations

Promulgated first under section 504 of the Rehabilitation Act,[79] and then under the Child Abuse Amendments,[80] the Baby Doe regulations set very stringent limits for decisions with respect to handicapped newborns. The regulations base decision making about care almost entirely on the likelihood of whether treatment will contribute to the survival of the impaired newborn. The regulations absolutely prohibit any consideration of likely quality of life for the infant.

Under the regulations, there are only three circumstances in which care may be withheld from a handicapped newborn:

(1) If, in reasonable medical judgment, the child is chronically and irreversibly comatose;[81]

(2) If, in reasonable medical judgment, the care would only prolong dying—that is, it would not be effective in ameliorating all of an infant's life-threatening conditions or otherwise would be futile in terms of survival;[82] or

---

77. Raymond S. Duff and A.G.M. Campbell, *Moral and Ethical Dilemmas in the Special-Care Nursery*, 289 NEW ENG. J. MED. 890, 892-94 (1973).

78. For a description of the context, see MEISEL, *supra* note 53, § 14.6–.7.

79. Rehabilitation Act of 1973, Pub. L. No. 93-112, 87 Stat. 394 (1973) (codified as amended at 29 U.S.C.A. § 794 (Supp. 1992)).

80. Child Abuse Amendments of 1984, Pub. L. No. 98-457, 98 Stat. 1749-64 (1984) (codified in scattered sections of title 29 U.S.C.A). The regulations were originally promulgated under § 504 of the Rehabilitation Act in response to two highly publicized cases of withholding of medical treatment for handicapped newborns. *See* Bowen v. American Hosp. Ass'n, 476 U.S. 610, 617-23 (1986). The Supreme Court held that the regulations had been improperly promulgated under the Rehabilitation Act. *Id.* at 647. A second round of "Baby Doe" regulations, adopted under the Child Abuse Amendments, require states to protect infants with life-threatening conditions from "medical neglect" in exchange for receiving child abuse prevention funding. *See* 45 C.F.R. § 1340.1–.20 & app. (1991).

81. 45 C.F.R. § 1340.15(b)(2)(i) (1991).

82. *Id.* § 1340.15(b)(2)(ii).

(3) If, in reasonable medical judgment, the care would be virtually futile in terms of survival and under the circumstances would be inhumane.[83]

In addition, there are no circumstances under which nutrition, hydration, or comfort care may be withheld, despite the infant's prognosis.[84]

These regulations impose remarkable limits on family decision making with respect to handicapped newborns. They express a commitment to the preservation of life in nearly all circumstances, which is clearly at odds with the liberal account sketched above. They do not allow parents to consider whether death is preferable to continued existence with truncated capacities, except in the case of chronic coma. Nor do the regulations allow parents to forego inhumane treatment, unless it would be virtually futile in terms of survival.

The Baby Doe regulations apply to states that choose to receive federal funding for their programs to prevent child abuse. In order to receive the funding, states are required to have statutory definitions of child abuse, including medical neglect, that roughly track the regulations.[85] A few states have chosen to forego the funding.[86]

3. *The Current Confusion*

Case law after the Baby Doe regulations is very limited. Three reported appellate decisions have involved parental requests to withdraw care.[87] In each case, the patient was an infant, chronically comatose, and had no likelihood of recovering any cognitive function. In all of the cases, the court permitted discontinuation of the care. In one case, the court specifically described the discontinuation as permissible because the infant was both irremediably comatose and "terminally ill," despite indications that the infant could live from one to five years with aggressive

---

83. *Id.* § 1340.15(b)(2)(iii).

84. *Id.* § 1340.15(b)(2) & app.

85. *Id.* § 1340.14(b).

86. *See* Terry J. Barnett, *Baby Doe: Nothing to Fear But Fear Itself*, 10 J. PERINATOLOGY 307, 310 (1990).

87. *In re* Guardianship of Barry, 445 So. 2d 365 (Fla. Dist. Ct. App. 1984); *In re* L.H.R., 321 S.E.2d 716 (Ga. 1984); *In re* P.V.W., 424 So. 2d 1015 (La. 1982).

supportive therapy.[88] Ia second case, the court concluded specifically that the state's "Baby Doe" law permitted the discontinuation of therapy in cases of irreversible coma.[89] Meisel has argued that these decisions are not limited to allowing parental discretion when the child is terminally ill and comatose and that the situation for children appears to be developing in parallel to the situation for adults.[90] The utter paucity of cases involving children, however, together with the quite explicit language in these cases about the child's comatose and terminal condition, suggest that Meisel's conclusion is unduly optimistic.

Indeed, state statutes also appear to leave parents and the health care profession in uncertainty about whether parents have the power to discontinue care in cases that do not meet the strict Baby Doe criteria (or their analogues for older children). Some states give parents the power to consent to health care for their children but are silent about whether this includes the power to make nontreatment decisions.[91] A number of states define child abuse or neglect generally to include the failure to provide needed medical care.[92] Some explicitly provide that religiously motivated failure to seek care is not criminal child abuse.[93] Yet none of these state child abuse statutes have dealt with whether nontreatment decisions in cases of medical disagreement or uncertainty should be viewed as the failure to provide needed care.

Five states have spelled out explicit limits on the power of parents to discontinue care.[94] One of these statutes, Minnesota's, specifically tracks the Baby Doe regulations.[95] Louisiana prohibits any decision to deprive a child of nutrition, hydration, oxygen, or comfort care with the intent to cause or allow the death of the

---

88. *Barry*, 445 So. 2d at 370-71.

89. *P.V.W.*, 424 So. 2d at 1021-22.

90. MEISEL, *supra* note 53, § 13.7–.8.

91. *E.g.*, CAL. CIV. CODE § 25.8 (West 1982); UTAH CODE ANN. § 76-14-5(4)(a) (1983) (malpractice statute).

92. *E.g.*, ALA. CODE § 26-14-1(2) (1986); ALASKA STAT. § 47.10.010(2)(B) (1990); UTAH CODE ANN. § 76-5-110(1)(d) (1990).

93. *E.g.*, ALA. CODE § 26-14-1(2) (1986); ARK. CODE ANN. § 5-27-221(c) (Michie 1987); CAL. PENAL CODE § 11165.2 (West 1992); MASS ANN. LAWS ch. 273, § 1 (Law. Co-op. 1992).

94. *See infra* notes 95-101 and accompanying text (discussing the five states' limits on parents' power to terminate care).

95. MINN. STAT. ANN. § 260.015(5) (West 1992).

child.[96] Louisiana also prohibits the intentional deprivation of care that is "necessary to attempt to save the life of the child in the opinion of a physician exercising competent medical judgment," with three exceptions: profound and irreversible coma, a condition that will be terminal despite "every appropriate medical treatment," or care with potential risks that outweigh the potential benefits for survival.[97] Louisiana is the only state with a specific provision for parents to execute directives to withdraw or withhold care for ill children when the child's condition falls within the statutory provisions.[98] Like the Baby Doe regulations, Louisiana's statute asserts a preference for the preservation of life except when all cognitive capacity has been lost. Two states, Rhode Island and Indiana, provide that parents may not withhold nutrition, medical treatment, or surgical intervention to a handicapped child if that care is generally provided to similarly situated children without handicaps.[99] Finally, Arizona requires parents to provide medically necessary treatment for their children but exempts care that is not necessary to save life or that will only prolong the process of dying.[100]

Another entirely uncharted area in health care decision making for children is the role of family members other than parents. Unlike the family consent statutes for adults, the statutes described above generally deal only with parents or guardians. Meisel reports that in practice when parents are not available to make decisions for their children, attending physicians turn to other available family members, much as is done for adults. But there is no legal authority for this practice, either in case law or statute, and Meisel cites none.[101]

V. CONCLUSION

Thus, the legal picture of the role of families in health care decision making is quite different for children than for adults. For

96. LA. CHILDREN'S CODE art. 1553 (West Supp. 1992).
97. *Id.* art. 1554(1)–(3).
98. *See id.* art. 1557.
99. IND. CODE ANN. § 31-6-4-3(f) (Burns Supp. 1992); R.I. GEN. LAWS § 40-11-3(b) (1990).
100. ARIZ. REV. STAT. ANN. § 36-2281 (1986).
101. MEISEL, *supra* note 53, § 13.3.

adults, the law is increasingly authorizing family members to act for incompetent relatives. For children, the law is far more diverse. Parents are required to provide necessary medical care for their children, and in some states, they are specifically authorized to consent to health care on their children's behalf, but what these provisions mean is unclear. Case-law development has been largely cut off in the wake of the right-to-life perspective of the Baby Doe regulations. Other family members are almost entirely left out of the legal picture.

Yet if the liberal view of health care decision making is of interest—and, whatever its merits, it is the predominant view in bioethics today—this legal picture is backwards. Within this liberal framework, the roles of families for adult patients are principally reporting and implementing the patient's own choices or advocating for the patient's best interests. The role of parents in making decisions for their children, however, may extend to decisions which, within limits, shape their children's futures. Yet, under current law, families are given greater latitude in decision making for incompetent adults than for their children. Perhaps this conclusion shows that the liberal picture itself is flawed. Or perhaps it shows that the intervention of the Baby Doe regulations has unfortunately truncated development of legal understanding of the authority of parents to make health care decisions for their children.

Medical Ethics and Family Affairs

*Jay A. Jacobson, M.D.**

I. INTRODUCTION

I will explore how doctors regard their patients' family members in the process of making medical decisions. Two ethical considerations would seem to define potential ethical problems as they pertain to family members. These are physicians' obligations or duties toward family members and physicians' obligations of confidentiality to their patients. I will discuss ways in which physicians are guided or not guided with regard to these concerns and also speculate on how they behave and what factors might influence that behavior. I will do so from both a normative and a descriptive approach.

II. A NORMATIVE APPROACH

Physicians are not taught and are generally unfamiliar with either the Hippocratic Oath or the Codes of Conduct or Ethical Principles of the American Medical Association, sources that are frequently cited by nonphysicians as declaring the written norms of medical practice. Nevertheless, it is worth examining those norms to see what they prescribe regarding physicians and how they relate to their patients' families.

The Hippocratic Oath counsels: "[w]hatever houses I may visit, I will come for the benefit of the sick, remaining free of all intentional injustice, of all mischief and in particular of sexual relations with both female and male persons, be they free or slaves."[1] It is interesting that this prohibition on sexual activity does not distinguish between members of the patient's family and

---

* Chief, Division of Medical Ethics, LDS Hospital, Salt Lake City, Utah, and Professor of Medicine, University of Utah School of Medicine.

1. HIPPOCRATIC OATH, *translated in* Ludwig Edelstein, *The Hippocratic Oath: Text, Translation and Interpretation, in* ANCIENT MEDICINE: SELECTED PAPERS OF LUDWIG EDELSTEIN 3 (Owsei Temkin & C. Lilian Temkin eds., 1967), *reprinted in* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 329-30 (2d ed. 1983) [hereinafter HIPPOCRATIC OATH].

others, or even between the patient and members of the family—it appears to be quite a broad prohibition. However, one could also read this same text as defining the obligation of the physician as that of serving exclusively the sick individual in the household.

The oath further provides: "[w]hat I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself holding such things shameful to be spoken about."[2] Here, too, family members are not mentioned specifically. One might assume that the prohibition on spreading information abroad does not preclude sharing that information with members of the patient's family, but no such exception is stated.

The AMA Code of Conduct of 1847 ("1847 Code") divides physician obligations into three general categories: (1) duties of physicians to their patients; (2) duties of physicians to each other and to the profession at large; and (3) duties of the profession to the public.[3] Although no general section is devoted to the duties of physicians to patients' families, we can find some hint of an obligation to family members in the section concerned with duties of the profession to the public. In that section, the 1847 Code provides:

> [A]s good citizens, it is the duty of physicians to be ever vigilant for the welfare of the community, and to bear their part in sustaining its institutions and burdens: they should also be ever ready to give counsel to the public in relation to matters especially appertaining to their profession, as on subjects of medical police, public hygiene, and legal medicine. It is their province to enlighten the public in regard to quarantine regulations . . . and in regard to measures for the prevention of epidemic and contagious diseases . . . .[4]

Quarantine practice at the time, however, consisted of keeping a presumably infectious patient at home, thereby protecting the community, but placing the family at increased risk.

---

2. *Id.* at 330.

3. AMERICAN MEDICAL ASSOCIATION, CODE OF MEDICAL ETHICS (1847), *reprinted in* ETHICS IN MEDICINE: HISTORICAL PERSPECTIVES AND CONTEMPORARY CONCERNS 29-34 (Stanley J. Reiser et al. eds., 1977) [hereinafter AMA CODE OF ETHICS].

4. *Id.* at 33.

Regarding confidentiality, we find in the section called Duties of Physicians to Their Patients:

> Secrecy and delicacy, when required by peculiar circumstances, should be strictly observed; and the familiar and confidential intercourse to which physicians are admitted in their professional visits, should be used with discretion, and with the most scrupulous regard to fidelity and honor. The obligation of secrecy extends beyond the period of professional services;—none of the privacies of personal and domestic life, no infirmity of disposition or flaw of character observed during professional attendance, should ever be divulged by him except when he is imperatively required to do so. The force and necessity of this obligation are indeed so great, that professional men have, under certain circumstances, been protected in their observance of secrecy, by courts of justice.[5]

We see in this, as in the Hippocratic Oath, that it is not necessarily the case that all things that physicians learn should be kept secret from all other individuals. For example, the nature of an illness per se does not seem to fall under the provisions of secrecy or confidentiality. Instead, the "privacies of personal and domestic life, the infirmity of disposition or the flaw of character" would seem to be the "secrets" not to be discussed. Ironically, it seems that these items are most likely to be known by members of the patient's family anyway, whereas the details or the nature of the patient's illness would not necessarily be known by this intimate group.

With respect to information about the nature of a patient's illness, it seems clear in these codes that particular kinds of illness should be handled in certain ways. For example, the 1847 Code states:

> A physician should not be forward to make gloomy prognostications, because they savour of empiricism, by magnifying the importance of his services in the treatment or cure of the disease. But he should not fail, on proper occasions, to give to the friends of the patient timely notice of danger, when it

5. *Id.* at 29.

> really occurs; and even to the patient himself, if absolutely necessary.[6]

Thus, if the news is bad, it seems it is to be preferentially shared with "friends of the patient," rather than the patient himself. It is curious to me that in this section, and in most of the remainder of the 1847 Code, the term "friends of the patient" appears where I might have anticipated "family." Perhaps it is meant to mean those who have the patient's best interest in mind, rather than the distinction that comes to mind today between acquaintances and relatives. Even with this broad interpretation, however, I read this not as an obligation to share information with the family, but rather to spare the patient from burdensome bad news. Good news, one might infer, could be shared with the patient and maybe even the patient's family.

In the current Principles of Medical Ethics, drafted in 1980, which reduce the rather lengthy 1847 Code to seven discrete principles, we see new and considerable attention to law as a guide for physician behavior.[7] Once again, though, the word family does not appear. One can still read an implied narrow focus on the individual patient but, interestingly, in the first principle, the physician's obligation appears to be a bit broader than we have seen previously: "A physician shall be dedicated to providing competent medical service with compassion and respect for human dignity."[8] Note that the recipient of that service is not defined or identified. We find here also a new obligation which appears to be at least as broad, if not broader, than the duties of public health first delineated in the 1847 Code: "A physician shall recognize a responsibility to participate in activities contributing to an improved community."[9]

With respect to confidentiality, we see the information to be treated as confidential further defined, at least by implication, and the limits of that confidentiality defined by the law. "A physician shall respect the rights of patients, of colleagues, and of other

---

6. *Id.*

7. *See* AMERICAN MEDICAL ASSOCIATION, PRINCIPLES OF MEDICAL ETHICS (1980), *reprinted in* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 331-32 (2d ed. 1983) [hereinafter AMA PRINCIPLES OF MEDICAL ETHICS].

8. *Id.* at 331.

9. *Id.* at 332.

health professionals, and shall safeguard patient confidences within the constraints of the law."[10] Here one might imagine that that which is to be held confidential would be the verbal information the patient gives to the physician, in part because the patient assumes that it will not be shared with others. It is not as clear that medical diagnoses based on physical examination or laboratory results might enjoy this same protection. Furthermore, it appears that, rather than making the claim that holding information confidential is likely to enjoy the protection of the law, as did 1847 Code, here the inference is that the law may in fact force release of such information. Not only the law, but third-party payors, may require the physician to report "confidential" information despite the explicit objection of patients. The fate of that information quickly passes beyond the physician's control.

Insurance companies may forward detailed copies of bills to patients' spouses or parents which alert them to procedures or diagnoses that the patient might otherwise have concealed from them. State health departments require physicians to report communicable diseases. Some health departments do contact notification in the case of some sexually transmitted diseases. This has the effect of sometimes informing individuals of their spouse's infection, but it also may disclose a de facto case of marital infidelity—something Hippocrates might have kept to himself, "holding such things shameful to be spoken about."

None of this is meant to imply that doctors always conform to the requirements of law or insurers where confidentiality is at stake, but their behavior has not been well studied. We know that they will use deception or conceal patient information to reduce patient expenses, perhaps they would do so to avoid patient embarrassment or marital discord as well.[11] Extra-medical circumstances may be germane here, such as how closely the doctor identifies with the patient and the doctor's views on infidelity and candor in marriage. Patients often conspire with physicians to keep potentially problematic information from their insurance carriers. They pay cash for a visit and do not file a claim. Some patients might prevail on a physician to not report a

10. *Id.* at 331.

11. *See* Dennis H. Novack et al., *Physicians' Attitudes Toward Using Deception to Resolve Difficult Ethical Problems*, 261 JAMA 2980, 2981-82 (1989).

communicable disease or not do a diagnostic test which, if positive, would be reported directly from the laboratory and eventually get reported to their spouse.

III. A DESCRIPTIVE APPROACH

If an examination of oaths and codes leaves us a bit disappointed in the search for principles or norms that guide physicians in the way they should regard members of their patient's family, perhaps a descriptive approach could be more fruitful. This might include an account of what physicians in training observe their mentors doing and also what is done by practitioners.

With respect to the former, it might be useful to comment briefly on the medical school curriculum. Each class entering medical school attends lectures as a group for the first two years. During that time, their faculty consists mostly of non-physicians. The subject of study is the human body, not the patient. Year three, however, is clinically oriented. Role models and teachers are physicians, but they are differentiated, that is to say, they are specialists. Also, they generally teach without specific involvement of other specialists.

Thus, students first learn a doctor/body relationship and then they learn a doctor/patient relationship not generically, but specifically. They learn, for example, a surgeon/patient, internist/patient, psychiatrist/patient relationship. While each of these seems to have a definable set of parameters, it is less clear to me whether the medical specialty defines the doctor/patient/family relationship in a particular way; but it well may for several reasons, such as the age of the patients seen or the type of medical problems they present. Before we examine these, though, I'd like to turn for a moment to the way in which students learn to gather and organize data about patients.

As students learn the practice of medicine, rather than the medical science that undergirds it, the word family occurs infrequently but regularly. Students must learn how to take a medical history and perform a physical examination. There is an order and a ritual manner for conducting these and recording them. The student first elicits a personal medical history, ideally from the patient him or herself. Next comes the physical examina-

tion, followed by a review of systems. Then comes something called the family history, and finally a social history.

The family history has as its focus heritable diseases and tends to have a vertical orientation with a special emphasis on the set of illnesses in family members who preceded the patient and, when appropriate, those that are or were descendant from him or her. A typical inquiry might be, does or did anyone in your family have diabetes or heart disease? The purpose of the inquiry is not so much to ask after the health of the patient's mother or father out of courtesy or concern, but rather out of curiosity for details that might pertain directly to the health of the patient. No notion of patient confidentiality accompanies this line of questioning. If the patient responded, "I'd rather not say, my mother considers her health quite personal," I'm sure doctors would be shocked. I have never heard of such a response. It is expected that physicians will ask patients questions about the health and, where applicable, the cause of death of their relatives, and that patients will answer to the best of their ability. Physicians will often construct rudimentary family trees and insert them into the record, indicating which relatives are living or dead and which are affected by some inheritable disease or condition. There is a presumption that patients will know, in considerable detail, the health of their relatives. It doesn't necessarily follow that they learned the facts from their relative's doctors, but that seems plausible. They may have learned about them from the relative him or herself or some other family member. If this is so, it suggests that family members do not hold health matters "shameful to he spoken about" to each other. Perhaps this provides some justification for the Hippocratic physician who chooses to spread information abroad, but not beyond the family circle.

The social history, on the other hand, seems a bit more horizontal. It branches laterally from the patient and adds detail to what we know about him or her in the present moment. Is he married? Does he work? At what job? Does he drink or smoke? The questions asked in this section seem to have more than one theme. One is, again, disease risk factors for and clues to disease from our patient, but we also seem to learn a bit more about him or her. Why do we want to know if he is married? It could be procedural or economic: Who is responsible for the bills and how likely is it that the bill will be paid? It could also be anticipatory:

Whom should be notified in case of death, or consulted in case of subsequent incompetency.

For the practitioner, oriented to families in ways defined by the medical curriculum and the standard medical interview, the age of the patient may still turn out to be quite an important factor with respect to attitudes toward family members. Consider pediatrics or, particularly, neonatology. With such young patients, does a physician really have a doctor-patient relationship? There might be an obligation to the patient, but if form follows function, obligation may follow, or at least be influenced by, relationships. Consider the problem of doing what is best for the patient in the pediatric context. Recall the Hippocratic Oath that makes this a sole imperative. Then consider who we think knows what is best for individuals of this age. Surely it is their parents or even more specifically, their mother. Mothers speak for their children, decide for them and, in a direct or indirect way, pay for their health care. No wonder then that a doctor treating a child or infant patient seems to respond to a different set of ethical imperatives than one treating an adult patient. It would appear strange if the doctor acknowledged no obligation to the parent and felt obliged to keep the nature of the child's illness secret from them.

Pediatricians seem to invert the usual Hippocratic notions. They work, in part at least, for the benefit of the parents and they share extensively the details of the child's condition with them. As children grow older, however, the way pediatricians treat them and members of their family begins to change. The mother now, instead of being asked to stay for her adolescent daughter's examination, is asked to leave. Physical findings, lab tests, and patient confidences, especially if they all pertain to pregnancy, seem to garner a level of confidentiality quite different than what these same physicians accord to their younger patients. Here, too, the law has had considerable influence on what physicians do and regard as ethical practice.

At the other end of the age spectrum, I see that physicians pay particular attention to members of their older patient's family, sometimes to the point of almost ignoring the patient. As they take a medical history, for example, doctors often face and address questions to family members who are encouraged to enter the examining room with the patient. There may be patient-centered reasons for following this practice, such as pronounced disorders

of hearing or speaking or serious memory loss or dementia, but the practice is prevalent even in their absence. Doctors with elderly patients often invite or allow family members to remain in the room during a complete physical examination.

Aside from age, a medical specialty determines what kinds of problems come to a physician's attention. Obvious examples are the specialties that deal with sexual or reproductive health problems. A physician who sees male patients with impotence or one who sees women concerned about infertility are both more likely to personally confront questions of family involvement—more often than a dermatologist might.

Another type of medical or ethical problem that seems to shape doctors' regard for family members is the incompetence of the patient or the specter of death or irreversible illness—especially when these occur together. It should be obvious to you that, in the case of incompetent patients, doctors feel obliged to obtain permission for treatment from the patient's family members. In this regard, doctors do not tend to follow a legally established hierarchy for decision making, but tend to consult available family members, especially those who choose to accompany or visit the patient.

If there is no apparent intrafamilial disagreement with a self-proclaimed family spokesperson, physicians are likely to follow the family member's recommendation about treatment. If there is disagreement, physicians are likely to follow the course which looks least treacherous in terms of litigation.[12] That might mean that they would continue life-sustaining care in some cases that a family member regarded as futile or that they would postpone elective surgery in some cases where a family member thought it desirable.

What may be a bit more surprising is that, in the situation of competent patients with serious and/or irreversible illnesses, physicians show a preference for discussing possible limitations of treatment with family members rather than the patient. The best studied example is that of the "do not resuscitate" order. Such

12. *Cf.* Henry S. Perkins et al., *Impact of Legal Liability, Family Wishes, and Other "External Factors" on Physicians' Life-Support Decisions*, 89 AM. J. MED. 185, 187-89 (1990) (physicians' management choices regarding life-support treatment influenced by assumption of legal liability or legal immunity).

orders are infrequently addressed with patients,[13] even though many hospitals have policies stating that this should be done.

The last situation I will mention is the death of a patient. Here it seems clear that it is the patient's family that has an acute need for care and compassion and not the patient. Medical texts and codes have little to say about the doctor's obligation to provide these. Until recently it was common for doctors to practice deception as a way of meeting this perceived need. They would telephone the family or have someone call to say that the patient had changed for the worse and it would be best for the family to be present as soon as possible. The physician would meet the family, explain in person that the patient had died, discuss the circumstances, and offer what comfort he or she could. Recently, the practice has changed to some extent. In teaching hospitals, notification of death often is a task delegated to the resident and, regardless of whose responsibility it is, the process is often done over the phone.[14]

Death provides one last way of analyzing physicians' concerns with a patient's family: the physician's acknowledgement of the death outside of the medical context through a written note or attendance at the patient's funeral. Physicians rarely respond to a patient's death and to the patient's family in either of these ways.[15] This does vary by medical specialty, but, in general, doctors perform these functions much less often than dentists do.[16]

IV. CONCLUSION

In summary, I would like to restate that there is a paucity of written material within medicine that defines the duty of the

---

13. Susanna E. Bedell & Thomas L. Delbanco, *Choices About Cardiopulmonary Resuscitation in the Hospital: When Do Physicians Talk with Patients?* 310 NEW ENG. J. MED. 1089, 1090-91 (1984).

14. Susan W. Tolle et al., *Physician Attitudes and Practices at the Time of Patient Death*, 144 ARCHIVES INTERNAL MED. 2389, 2389-40 (1984); *But see* Terri A. Schmidt & Susan W. Tolle, *Emergency Physicians' Responses to Families Following Patient Death*, 19 ANNALS EMERGENCY MED. 125, 127 (1990) (minority of emergency physicians notify family of death by telephone).

15. Schmidt & Tolle, *supra* note 14, at 127; Tolle et al., *supra* note 14, at 2390.

16. *See* Gary T. Chiodo & Susan W. Tolle, *Patient Death and Bereavement: What is the Dentist's Role?*, 8 SPECIAL CARE DENTISTRY 198, 199 (1988).

generic physician to his or her patient's family. I have drawn on personal experience and a few available studies to describe how doctors behave with respect to their patients' families. I have explored the issue from the standpoint of medical norms, medical education, patient circumstances, and medical specialty. I intentionally avoided the specialty called family practice. I do not want to end with the suggestion that physicians in this field are not different from their colleagues, but rather that, as a hospital-based physician, I am much less familiar with how they may differ. For the rest of medicine, however, I think that major ethical obligations, perceived and performed, replicate those outlined in the 1847 Code, with a slight modification: (1) duties of the physician to the patient; (2) duties of physicians to each other; and now (3) duties of physicians to third-party payors.

Telling Medical Stories: Sharing Information Among Doctors, Patients, and Families*

*Martha Minow***

Dr. Jay Katz recounts this story:

> The senior physician asked the intern how much he knew about "patients as human beings." The question led to a rather nonproductive exchange that the intern ended abruptly with the exasperated comment: "I cannot answer your questions. You're interested in patients, I'm interested in the disease in the body in the bed."[1]

Dr. Stephen Hoffmann, in telling the story of his own medical training, revealed an attitude that Dr. Katz would have preferred: "With a magnanimity that borders on the incredible, patients treat us to themselves—that is, if we're lucky enough to really hear them."[2]

I would like to consider the stories doctors, patients, and families tell to and about one another as part of an inquiry into the ethics of medical encounters. My most basic question asks how doctors should communicate technical and especially statistical information, and whether they be obliged to ensure that patients and their families understand that information. I also would like to use the opportunities provided by this question and this conference to reflect upon the contrast between information and stories as methods for making meaning about living, illness, risk, and dying.

Both questions call for attention to the problem of translation. How can and how should doctors, patients, and family members

---

** Professor of Law, Harvard University. Thanks to Joe Singer, Vicky Spelman, Ben Wizner, Judy Edersheim, Scott Brewer, Linda Emanuel, Jo Minow, and Newton Minow for their helpful suggestions.

1. JAY KATZ, THE SILENT WORLD OF DOCTOR AND PATIENT xix (1984).

2. STEPHEN A. HOFFMANN, UNDER THE ETHER DOME: A PHYSICIAN'S APPRENTICESHIP AT MASSACHUSETTS GENERAL HOSPITAL 55 (1986). Hoffmann continued that the physician's job ultimately is bearing witness to the lives and sufferings of patients. *See id.* at 59.

learn to translate from the language they use to make sense of what they see? Can they retain the meanings they understand while coming to comprehend the unfamiliar and potentially frightening information offered by someone in a different position?[3]

My inquiry arises at a time of threatened and actual legal regulation of doctor-patient communications. Despite constitutional objections, the Supreme Court of the United States has approved restrictions on what doctors may tell patients when the medical care occurs in a federally-funded family planning clinic.[4] At the same time, a public debate has arisen over a psychiatrist's decision to share with a biographer the tapes of his therapy sessions with a patient, thus allowing public access to private communications with that patient.[5] Also hotly contested are proposals to require

---

3. Hoffmann notes that the doctor must learn to translate; the "patient explain[s] his problem in the vocabulary of his own making," and the doctor "then attempts to capture the content in the strict language of medicine, losing as little as possible in translation." *Id.* at 169. In addition, "language itself, not the stethoscope or blood pressure cuff, is the major instrument through which doctors work. It is through a patient's language, a medium both infinitely expressive and infinitely treacherous (because it is so subject to interpretation), that a physician arrives at a diagnosis and plan." *Id.* at 170.

4. Rust v. Sullivan, 111 S. Ct. 1759 (1991) (rejecting constitutional challenge to regulation forbidding recipients of family planning funds under Title X of the Public Health Service Act from advocating abortion as an option in family planning). As many have noted, the holding that approves restriction on physician speech may be confined to the context of abortion; it also seems limited to the context of a condition on public expenditures. But, as one observer noted, the case potentially poses a broad threat to physician-patient communications by raising as questions:

> How much control over conversations between doctors and patients can the federal government now claim for health care it funds through Medicare and Medicaid? To help control the costs of health care, could HHS [the Department of Health and Human Services] limit the information physicians could give such patients about alternative treatments? Under a system of national health insurance, could we have "state medicine," with the content of the doctor-patient dialogue prescribed by federal regulations, at least as long as some private physicians were available for those who could pay for them?

George J. Annas, *Restricting Doctor-Patient Conversations in Federally Funded Clinics*, 325 NEW ENG. J. MED. 362, 364 (1991).

5. *See* Alan A. Stone, *Confidentiality in Psychotherapy—The Case of Anna Sexton*, 325 NEW ENG. J. MED. 1450 (1991) (reviewing DIANE W. MIDDLEBROOK, ANNE SEXTON: A BIOGRAPHY (1991)). The psychiatrist's decision was authorized by the patient's literary executor who is also her daughter. *See generally* Tamar R. Lehrich, *To Bedlam and Part Way Back: Anne Sexton, Her Therapy Tapes, and the Meaning of Privacy*, 2 UCLA WOMEN'S L.J. (forthcoming 1992).

medical professionals to disclose their own health status in relation to HIV and AIDS.[6] Thus, public control of medical conversations has become a plausible and intense subject for policy makers and lawyers, a subject that puts at stake matters of individual dignity, equality, and fairness. It is these stakes I mean to acknowledge, although I will not be addressing these particular hot policy debates and legal questions.

Instead, my inquiry is explicitly ethical. For the common, daily versions of medical conversations I mean to address, law is rather peripheral. I am reminded of the Australian judge who wrote that the law marches with medicine "'but in the rear and limping a little.'"[7] Law will be late and largely crude in response to the delicate questions of trust in communication. The project here more directly deals with matters of ethical practice than with legal requirements, although the danger that law could get in the way of ethical practice must also be addressed. The special problems posed by the involvement of families as well as patients in communicating and sharing decision making with doctors inevitably raise issues of legal authority as well as ethical judgment. Cognizant of the legal context provided by both informed consent law and family law, let us pursue what ethical concerns should guide medical communications and what ongoing ethical inquiries should become part of medical practice.

I take as touchstones the physician's commitment to "do no harm";[8] the Golden Rule's ethical prescription not to do to another what you would not want done to you;[9] and this advice from one doctor to his trainees: "[W]henever the physician asked a patient to hold his breath, he should always do the same, so that he would

---

6. Peter Bayer, *AIDS Spawn Epidemics of Death and Fear*, S.F. CHRON., Aug. 5, 1991, at A15; Marlene Cimons, *U.S. Seen Urging Voluntary AIDS Tests for Surgeons*, L.A. TIMES, July 15, 1991, at A1.

7. ZELMAN COWEN, REFLECTIONS ON MEDICINE, BIOTECHNOLOGY AND THE LAW 5 (1986) (quoting Mount Isa Mines, Ltd. v. Pusey, 4 S.A. St. R. 88 (1971)).

8. *See* Curley Bonds, *The Hippocratic Oath: A Basis for Modern Ethical Standards*, 264 JAMA 2311 (1990); Michael R. Flick, *The Due Process of Dying*, 79 CAL. L. REV. 1121, 1125 (1991).

9. *See* DAVID WALLECHINSKY & IRVING WALLACE, THE PEOPLE'S ALMANAC 1314-15 (1975) (citing versions of the Golden Rule from Christianity, Judaism, Brahmanism, Buddhism, Confucianism, Taoism, Zoroastrianism, Islam, and a dissenter). *See also* KATZ, *supra* note 1, at 229 (relying on the Golden Rule to guide doctor-patient communications).

be aware of when the patient needed another breath."[10] How can the communications in these settings reflect these ethical commitments, given the differences in power and vulnerability manifested in the encounters among medical personnel, patients, and families?[11]

I start by addressing the problems posed by the patient's vulnerability, even after the adoption of the informed consent requirement. I then turn to difficult questions regarding how much doctors should tell patients. Next, I treat the further complications posed by the presence and involvement of family members. I recommend that good ethical practice by doctors should involve conveying statistical and other technical information in a manner that the patient and patient's family actually understand, and I offer some modest suggestions toward that end.

Finally, I argue that the ethical doctor must attempt to see and communicate with patients as individuals and human beings—that this is the first step toward genuinely shared understanding. I use the notion of storytelling throughout in an effort to highlight the meanings of medical encounters in the lives of both patients and doctors. Through narrations, we each try to organize and make sense of our experiences, and thus do more than simply exchange or withhold information.

I. How Much To Tell

If at one time physicians chose to withhold information—or lie—in dealings with patients, that tradition occurred before "paternalism" became a term of criticism. It also occurred before medicine became a world of specialties, a world in which patients deal often with teams of physicians or rotating clinical staff members instead of a family doctor, and a world in which malpractice lawsuits permeate the consciousness of the medical community.[12] Perhaps paradoxically, both the increasing array of

---

10. Hoffmann, *supra* note 2, at 45 (paraphrasing advice given by Dr. Plotcyk, Hoffmann's professor at Harvard Medical School).

11. The critical roles of nurses and other medical personnel to the care and treatment of patients and families deserve greater attention than I can give them here. For a discussion of some of these issues, see Margaret Mahon, *The Nurse's Role in Treatment Decisionmaking for the Child with Diabetes*, 6 Issues L. & Med. 247 (1990).

12. *See generally* Katz, *supra* note 1, at 22-23 (noting context for the adoption of

medical options and the continuing presence of medical uncertainty fueled the shift to the regime of informed consent, under which doctors are directed to inform patients of their medical status and options and to obtain consent before treatment.

Prior to and now following the regime of informed consent, however, risks of troubled communication remain. One risk is that physicians—and patients—will not tell each other enough to ensure that each will understand the other and contribute meaningfully to decisions. Another risk is that physicians will communicate too technically, in a fashion that the patients cannot comprehend or use meaningfully. Behind each of these risks is the larger danger that the meaning of medical encounters may get lost in the exchanges of both silence and information.

A. Why Doctors—and Even Patients—Do Not Tell Enough

Historically, doctors justified practices departing from full disclosure on the view that telling a patient diagnostic information or information about a potential treatment could do more harm than good. For example, Dr. Jay Katz tells the story of a distinguished French nephrologist who had examined a peasant

> who lived some 40 miles outside of Paris and suffered from chronic renal failure. The condition was a rapidly progressive one and would soon lead to death unless the patient were placed on renal dialysis. Yet the patient was not offered this option. Instead, he was told that no medical treatment existed that would help him. When I asked the nephrologist why he had concealed the alternative of dialysis from his patient, he reacted with surprise, as if the answer were self-evident: "To say more would have been cruel. Peasants do not adjust well to a permanent move to a large city." Dialysis would have required a permanent relocation.[13]

Based on this judgment, the doctor justified his nondisclosure of the treatment option.

---

informed consent rules in the 1957 and 1980 versions of the Principles of Medical Ethics).

13. *Id.* at 5.

Sometimes, doctors have justified refraining from telling a patient information, such as a particularly serious diagnosis of imminent death, on the theory that the knowledge itself would induce more suffering for the patient.[14] Other doctors have defended decisions to withhold the fact that a patient is participating in the control group of a research investigation. In one such instance, the doctors explained that they sought to spare parents the burden of consenting to participation in an experiment where their children would simply receive conventional rather than experimental treatment.[15]

In each of these contexts, the physicians presumed that they knew what was best for the patient or patient's family; they also presumed that knowledge itself could be harmful or painful. While a physician at times may know something about the individuals involved, these physicians did not proffer specific evidence that knowledge itself would be disturbing to the patient. At the same time, withholding information can have the effect of reinforcing the physician's power and authority and also shielding the physician from irritating demands, questions, or resistance,[16] thereby casting doubt on the physician's avowed intent of sparing the patient pain.

The doctrine of informed consent is predicated on the conclusion that protecting the patient from knowledge is not

---

14. If this involves lying, the doctor may end up undermining the trust of either this patient or future patients. In a 1909 essay, Richard C. Cabot noted that lying to the patient may save the patient some suffering, but "[h]is wife has now acquired, if she did not have it already, a knowledge of the circumstances under which doctors think it merciful and useful to lie. She will be sick herself some day, and when the doctors tell her that she is not seriously ill, is she likely to believe them?" *Id.* at 25-26 (quoting Richard C. Cabot, *The Use of Truth and Falsehood in Medicine: An Experimental Study*, 5 AM. MED. 344 (1903)).

15. Robert D. Truog, *Randomized Controlled Trials: Lessons from ECMO*, 40 CLINICAL RES. 519, 522 (1992). This study used the technique of "randomized consent" by which eligible subjects are randomly assigned to the experimental and conventional therapy, but informed consent is sought only after this randomization and only for those assigned to the experimental treatment. *Id.* at 521; Marvin Zelen, *A New Design for Randomized Clinical Trials*, 300 NEW ENG. J. MED. 1242 (1979). The defense for this practice is that no special disclosure or consent would be needed for the conventional treatment; the practice also shields the investigators from patient or family pressure to assign an individual to the experimental treatment. Truog, *supra*, at 522; Zelen, *supra*, at 1245.

16. *See generally* KATZ, *supra* note 1, at 132-141 (exploring heart transplant surgeon's failure to disclose doubts to patients and resulting freedom from anything but patient deference).

usually justifiable. The autonomy and dignity of the patient require sufficient involvement and access to information about the medical questions to enable the patient to make relevant decisions about treatment and about life more generally. Imagining reasons unrelated to the patient's own personality and experience for withholding information—even reasons predicated on what the doctor would prefer if he or she were the patient—is not consonant with the doctrine of informed consent. Instead, the doctrine implies that the doctor's incentives to avoid disclosing information interfere with the patient's good itself when conceived in terms of autonomy and dignity. The same issues arise as the doctor considers not only what information to disclose but also how to contrast or interpret the information; this, too, should be a joint venture, involving both doctor and patient.

Dr. Jay Katz argues further that doctors should not accept prematurely a patient's refusal to engage in conversations about fears, misconceptions, or doubts about diagnosis or treatment; instead the doctor should press the patient to face these matters.[17] As his argument sensitively explores, however, the doctor's position of authority and greater knowledge may continue to undermine communication with patients. Moreover, the knowledge and power of the doctor affect the patient's understanding and

---

17. *Id.* at 162. Sometimes doctors justify not telling on the theory that the patient does not want to know. Surgeon Richard Selzer's statement illustrates this attitude:

> The patient has a right to know everything. The patient may not wish to know everything. Ever since the 1960s, we've been told that the patient must be the colleague of his doctor and that the decision-making must be shared equally by the two of them. Intellectually and philosophically I certainly do agree. Unfortunately, it doesn't quite work out that way. When I try to call the patient in on a consultation and say, "Which alternative would you prefer?" *invariably* the patient says, "What do you mean, which alternative? I want you to tell me what to do. You're the doctor." The only unspoken word is daddy; tell me what to do, daddy. When a person is desperately ill or frightened, there is a certain kind of regression that makes you want to place yourself in someone's loving care. It is the responsibility of the doctor to have the courage to make decisions for the patient, in as kind and wise a way as he can. To be a father or mother and comfort. There are, of course, strong people who want to know everything and decide everything but they are a distinct minority.

Id. at 126 (quoting F. Middleton, *Profile: Richard Selzer*, 1 NEW HAVEN MAG. 37 (1983)). Katz in turn queries whether Selzer ever tried to explore the patient's response, or to emphasize that treatment choice properly is the patient's decision. Katz argues that the doctor needs an idea of patient's preferences and needs before making a recommendation and that doctors hear "daddy" too readily in patients' statements. *Id.*

hopes, which in turn may critically influence the patient's well-being.

Largely due to this power imbalance between doctors and patients, critics argue that practice departs from ideal in the informed consent regime. Some say that physicians intentionally or unintentionally use their position of authority to shape the disclosures in order to secure consent by patients.[18] Others argue that physicians reinforce certain social attitudes and conventions by cutting off conversations that challenge them.[19] There is also some evidence that physicians may offer poorer efforts at explanation to members of racial minorities.[20] Sometimes the economic and institutional contexts for delivering health care prevent doctors from spending time conversing with patients.

For whatever reason, many patients do not receive information that would be useful or even critical to their ability to make rational health care choices. One researcher who observed two gynecology settings reported:

> I never heard doctors explain to women, when prescribing the [birth control] pill, the possibility of temporary infertility if they stopped using this method. This was explained if women came to the doctors to stop the pill, but many women, like Maria, stop on their own. When such basic information is not forthcoming from doctors, the results can be devastating.[21]

Here, the problem of not telling may arise in part because the doctors do not listen to what patients want to tell about the social context for their birth control choice or their larger life plan.[22] The doctor communicates through tone and shifts in conversation what the patient should feel free to discuss—and what not to discuss.[23] The doctor without much reflection may assume that

---

18. *See id.* at 26; ALEXANDRA D. TODD, INTIMATE ADVERSARIES: CULTURAL CONFLICT BETWEEN DOCTORS AND WOMEN PATIENTS 48-49 (1989).

19. *See* HOWARD WAITZKIN, THE POLITICS OF MEDICAL ENCOUNTERS 81, 107-110, 141 (1991) (discussing how doctors reinforce conventional work, family, and gender roles by cutting off critical patient communications).

20. TODD, *supra* note 18, at 77.

21. *Id.* at 89-90.

22. *See id.* at 82-85, 90 (providing examples of women volunteering information they feel is pertinent and doctors not listening or shifting to discussion of biological processes severed from social events).

23. *See* WAITZKIN, *supra* note 19, at 131.

a patient's narrative has little to do with the medical assessment and so indicates this, even though the judgment may reflect a problem in translation more than a problem in relevance. Patients may not tell the doctor enough for other reasons. A patient may be nervous, frightened, or embarrassed.[24]

In this light, some of the failures by doctors to tell enough may stem from patients' failures—but again, the imbalance of authority between doctor and patient complicates matters. The patient's failure to ask questions or disclose information is likely to arise from the doctor's failure to set the patient at ease or to ask inviting questions. The problem can be pronounced even for the sensitive doctor: Dr. Hoffmann, who marveled at the gift of patients' stories, also reported that "[u]sually as doctors we do not withhold information so much as offer it in less explicit terms, trusting to patients to tell us if they want to know more."[25] Not all patients understand this implicit arrangement or feel comfortable asking for more.

Can doctors learn to imagine why a patient does not ask for more information? Can a doctor come to imagine the patient's experience as vividly as the doctor who learns when to direct a patient to breath by holding a breath while asking the patient to do so? Recent experiments in medical education aim in precisely this direction.[26]

Limiting how both doctors and patients understand one another is the fact that, as Jay Katz puts it, "even in their most intimate relationships, human beings remain strangers to one another" and perhaps remain strangers to oneself.[27] Yet even given these limits, doctors—and patients—can pay attention to the preconditions for hearing one another. A special challenge, though,

---

24. *See* TODD, *supra* note 18, at 90, 92 (discussing women patients too nervous to discuss reproductive issues).

25. HOFFMANN, *supra* note 2, at 139. One study indicates that the doctors asked several times as many questions as did the patients, and the doctors made about twice as many statements as did the patients.

26. *See* Lisa Belkin, *In Lessons on Empathy, Doctors Become Patients*, N.Y. TIMES, June 4, 1992, at A1. Short of literally experiencing what the patient experiences, doctors can incorporate into the obligations to inform and obtain consent the practice of really listening and asking specific questions about what the patient understands and does not understand.

27. KATZ, *supra* note 1, at xviii.

is posed by the use of information of a technical and especially statistical nature.

B. Telling Too Technically

Equal to the risk that doctors and patients will not tell each other enough is the danger that doctors and patients will not understand what the other tells. This danger is especially pronounced with technical and statistical information. I argue here that sensitive ethical practice of informed consent would engage doctors in the task of ensuring that the patient understands technical information, including statistical information. I do not mean to argue that doctors must ensure that patients understand, for that test is too difficult to administer or to meet, but doctors should take steps likely to promote that understanding.

This remains a tall order, especially when it involves statistical information about risks and benefits, because most people over emphasize risk or do not understand it.[28] Doctors' professional training typically involves such extensive drilling in risk assessments and they become so familiar with the use of quantitative information that they may not fully comprehend how people without such training hear statistical disclosures. Yet, the task is even more profoundly complex if doctors themselves do not comprehend the statistical information they offer patients—and there is evidence that this is sometimes the case.

1. Two Worlds

Like C.P. Snow's thesis that scientists and humanists inhabit two separate worlds,[29] let us explore the possibility that most doctors and most patients inhabit two separate worlds when it comes to technical and statistical information. On this theory, doctors do not comprehend what patients do not understand, and

---

28. *See generally* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases, in* JUDGMENT UNDER UNCERTAINTY 3 (Daniel Kahneman et al. eds., 1982) (discussing biases found in intuitive assessments of probabilities).

29. CHARLES P. SNOW, THE TWO CULTURES AND THE SCIENTIFIC REVOLUTION 2-6 (1959). The effect of local and national cultures on medical practice further complicates communication and patients' experiences. *See* Mary-Jo D. Good et al., *American Oncology and the Discourse on Hope*, 14 CULTURE, MED. & PSYCHIATRY 59, 59 (1990).

patients do not know how to ask questions that would produce answers they could comprehend. The problem is not merely that doctors have a kind of knowledge that patients do not have; doctors and patients may systematically diverge in how they make sense of the world, how they process information, how they narrate events, and even in what they value. Members of partially or fully distinct communities, doctors and patients need to approach one another with a recognition of a basic translation problem.

Accounts of medical education support this theory. For example, Dr. Hoffmann reported that during his first year of medical school,

> we learned a formula for everything. How do you express the flow of sodium ions into cells? How do membranes generate the electricity that flows at lightning speed down a nerve? How do notions as abstract as the pressures and volumes inside the heart translate into something as real as the flow of blood? . . . By deft acts of blackboard prestidigitation, our chalk-addicted professor would convert one commodity effortlessly to some other.[30]

During hospital rotations, Hoffmann learned of the trade-off time limits imposed: he could either talk informally with patients or spend more time reading their charts, and he concluded that chart-reading allowed him to identify and treat more problems even if it often reduced how much he could know the patients as people.[31]

Hoffmann implied but did not elaborate the effect of this trade-off: not knowing the patients as people could limit not only the quality of his own life but also his effectiveness and sensitivity in communicating with patients. It takes time to learn about a patient's points of reference, values, and ways of making sense of the medical experience. This kind of learning is possible but undervalued by the culture of medical institutions.

---

30. HOFFMANN, *supra* note 2, at 9. Hoffmann acknowledged that "[t]he body as an organism seemed to politely rebuff our reductionist attempts." *Id.*

31. *Id.* at 48. He found the pressures even more pronounced in the intensive care unit where he saw a competition over what would exert greater claim on physician time—the patients or their data. *Id.* at 68.

Whether due to professional training or discomfort with someone who is not a member of the profession, doctors may doubt their abilities to communicate technical information to patients. Jay Katz compared a conversation a surgeon had with a patient diagnosed with breast cancer and the conversation the same surgeon had with Katz, a fellow doctor, about the case. Katz found the surgeon able to explain to him but not to the patient the risks, indications, and contra-indications of alternate therapies.[32] The surgeon did not believe he could have had such a conversation with the patient, yet Dr. Katz noted that the surgeon used sufficiently nontechnical language in explaining the situation to him. This story could indicate that the doctor believed he and the patient inhabited more distinct worlds than they really did—or else that Dr. Katz was unable to assess whether a non-doctor could understand the surgeon's explanation, an alternative I find less plausible.

Yet, as evidence of the different world inhabited by patients, repeated studies indicate that patients often do not understand the informed consent forms they themselves sign.[33] Of course, these studies could instead be evidence of the stressful circumstances under which patients sign consent forms or the incomprehensibility of legal, rather than medical, technical information. Comparable studies indicate that juries understand little of jury instructions,[34] and consumers seldom comprehend loan disclosure statements.[35] Rather than two worlds—medical and nonmedical—people may inhabit many worlds, and these worlds are not separated simply by training or knowledge but more profoundly by cognitive and emotional styles. This conclusion, if anything, bolsters the basic point that patients often do not make sense of

---

32. KATZ, *supra* note 1, at 167-169.

33. HOFFMANN, *supra* note 2, at 129-30 (summarizing two *New England Journal of Medicine* studies regarding difficulty of following consent procedures).

34. Robert F. Forston, *Sense and Non-Sense: Jury Trial Communication*, 1975 B.Y.U. L. REV. 601, 606; Geoffrey P. Kramer & Dorean M. Koenig, *Do Jurors Understand Criminal Jury Instructions? Analyzing the Results of the Michigan Juror Comprehension Project*, 23 U. MICH. J.L. REF. 401, 401 (1989-90); Newton N. Minow & Fred H. Cate, *Who Is an Impartial Juror in an Age of Mass Media*, 40 AM. U. L. REV. 631, 662 (1991); Walter W. Steele, Jr. & Elizabeth G. Thornburg, *Jury Instructions: A Persistent Failure to Communicate*, 74 JUDICATURE 249, 249 (1991).

35. *See* Edward L. Rubin, *Legislative Methodology: Some Lessons from the Truth-in-Lending Act*, 80 GEO. L.J. 233, 236 (1991).

the very communications intended to inform them of the risks and benefits of proffered medical treatment.[36]

If this is the case, then genuine ethical concern for the patient should involve the doctor in rendering technical information comprehensible to the patient. Only then can concerns for the patient's dignity and autonomy, registered by the informed consent doctrine, be realized. Only then can the doctor and patient together construct the meaning of the patient's illness. But the obstacles here may include the doctor's own misunderstandings.

2. *Or One World—of Confusion?*

Doctors themselves may not understand the meaning of an illness, or the technical information they disclose to patients. Thus, problems of understanding may reach even deeper than the uncertainty that accompanies diagnostic and treatment judgments. There is evidence that statistical information, in particular, remains elusive and confusing to medical personnel.[37]

---

36. *See* HOFFMANN, *supra* note 2, at 129-130. Hoffmann states:

> [O]btaining consent demands a great deal intellectually from a patient. Deciphering technical terms, appreciating the role of drugs or tests, and conceptualizing unfamiliar procedures require a level of understanding that challenges even doctors who deal with the subjects every day. Asking, in addition, that patients not just understand the risks and benefits of medication, tests, or procedures, but be able to weigh them in the same way as doctors who have the benefit of years of training and practice seems unreasonable. How many patients (and how many doctors, for that matter) can weigh against the benefits of taking a single course of penicillin for a sore throat, for example, all the risks that lie in waiting? How can someone really make sense of, say, a 20 percent chance that his sore throat will be helped by penicillin versus a 1 percent chance that a rash will develop, an 0.005 percent probability that exfoliative dermatitis will be a side effect, or the 0.0003 percent chance that interstitial nephritis will occur?

Id. at 129. In addition, as Hoffmann notes, patients would have to be objective, and yet it is their bodies under discussion; this fact could alter their ability to make sense of the medical information. *Id.* Yet it seems a particularly American preoccupation to seek numbers to express medical information about diagnosis and prognosis. *See* Good et al., *supra* note 29, at 68.

37. One physician explained to me that it is often too hard to explain the data to patients and difficult even to understand it herself because studies produce conflicting assessments of risk. She also noted that statistical and risk/benefit analysis nonetheless occupy physicians in decisions about selecting staff and allocating resources as well as in direct patient care. Conversation with Dr. Karen Nelson, in Cambridge, Mass. (Feb. 29, 1992). The critical role of values in making sense of probabilistic information is illuminated by national differences in response to data about the treatment options for extremely premature infants. *See* Ernlé W.D. Young & David K. Stevenson, *Limiting*

The following episode illustrates this point. As part of a statistics seminar, a philosopher gave doctors and nurses a handout that read as follows:

> Assume that the following information is true.
>
> A new disease affecting people who read books about health care has recently been identified. It occurs in .1% of the population at risk. The symptoms are nasty, but the disease is not fatal. If it is left untreated, the victim has chills and fever for several weeks and suffers from moderate intermittent nausea and palsy. The symptoms then diminish and disappear. A new treatment eliminates these symptoms entirely, but it only works if administered prior to the onset of symptoms. That treatment involves no significant risk, and nearly always works. But it is costly, involves taking daily doses of a foul medicine, and, worse, requires total abstention from ice cream for eight weeks.
>
> A new test has been developed to identify victims of the disease presymptomatically. Its specificity and sensitivity are each 98%; that is, it identifies 98% of those who have the disease with a positive test result, and it identifies 98% of those who do not have the disease with a negative test result.
>
> The public health service has screened 100,000 potential victims of the disease with this test, in order to identify those who should be offered the treatment.
>
> I regret to inform you that your test was positive.[38]

The handout then asked the participants to answer two questions: "How likely do you think it is that you have the disease?" and "Do you want the treatment?"[39]

All but one of the doctors judged the likelihood to be at least 85%; 57% of the nurses said the odds were at least 85% while only

---

Treatment for Extremely Premature, Low-Birth-Weight Infants (500 to 750g), 144 AM. J. DISEASES CHILDREN 549, 550-551 (1990) (comparing contrasting approaches taken by Sweden, Britain, and the United States to the same treatment questions in light of risk/benefit assessments measured against different views of the value of spending shrinking resources).

38. SAMUEL GOROVITZ, DRAWING THE LINE: LIFE, DEATH, AND ETHICAL CHOICES IN AN AMERICAN HOSPITAL 139 (1991).

39. *Id.*

two said that the odds are not over 5%.[40] Indeed, the odds are not over 5%. As the philosopher explained,

> The incidence of the disease is one-tenth of 1% of the population. So of the 100,000 people who have been screened, 100 (approximately) have the disease. The test identifies 98% of the disease victims with a positive test result, so it will produce 98 positives—true positives—from those 100 people. The remaining 99,900 screened people do not have the disease, and the test will show that fact by providing most of them (98% of them) with a negative test result—a true negative.
>
> But the test isn't quite perfect; it will fail to identify 2% of those who are disease free as being disease free. It will give them a positive result—a false positive—even though they do not have the disease. Since there are 99,900 without the disease, that 2% rate of false positives will yield 1,998 more positive results, all of them false. The total number of positives based on the screening of 100,000 people will therefore be 1,998 plus 98, or 2096. But of those 2096 positives, only 98—4.7% of them—will be true positives! So on the basis of the data provided in the statement of the problem, the likelihood of a person in the population at risk actually having the disease, given a positive screening test result, is less than 5%.[41]

The range of answers by the doctors and nurses should give us pause about their abilities to communicate risks to patients.[42]

---

40. *Id.* at 141.

41. *Id.* at 142. In his book, *Innumeracy*, John A. Paulos works out a similar example about the chances you have cancer if the test is 98% accurate and the actual distribution is about .5% of the population. Then the conditional probability of having cancer, given that one tests positive, is only 49/248 or about 20%, if 10,000 tests are administered. JOHN A. PAULOS, INNUMERACY: MATHEMATICAL ILLITERACY AND ITS CONSEQUENCES 66 (1988).

42. Gorovitz notes that "[m]edical education does not produce statistical literacy, even though medical decisions are largely based on statistics." GOROVITZ, *supra* note 38, at 142.

3. Acknowledging and Addressing the Misunderstandings of Statistics

Patients and medical personnel alike may well overestimate risk, and in general have difficulty comprehending statistical information. Studies show that people respond differently to a problem involving probabilities depending upon how the question is phrased.[43] Because schools tend not to teach estimation, people lack a sense about the scope of numerical descriptions.[44] Statistical information about health risks can be especially misunderstood because the implications may be frightening for patients. Moreover, physicians need to think about the groups of patients over time while patients are bound to think individually about themselves.[45] As one person put it, the probability of a poor outcome may be 1/40 for the population, but if you are the one, the probability is 100%.[46]

---

43. *See* PAULOS, *supra* note 41, at 86-87. Paulos summarizes a study undertaken by Tversky and Kahneman:

> Imagine you are a general surrounded by an overwhelming enemy force which will wipe out your 600-man army unless you take one of the two available escape routes. Your intelligence officers explain that if you take the first route you will save 200 soldiers, whereas if you take the second route the probability is 1/3 that all 600 will make it and 2/3 that none will. Which route do you take?
>
> Most people (three out of four) choose the first route, since 200 lives can definitely be saved that way, whereas the probability is 2/3 that the second route will result in even more deaths.
>
> So far, so good. But what about the following? Again, you're a general faced with a decision between two escape routes. If you take the first one, you're told, 400 of your soldiers will die. If you choose the second route, the probability is 1/3 that none of the soldiers will die, and 2/3 that all 600 will die. Which route do you take?
>
> Most people (four out of five) faced with this choice opt for the second route, reasoning that the first route will lead to 400 deaths, while there's at least a probability of 1/3 that everyone will get out okay if they go for the second route.
>
> The two questions are identical, of course, and the differing responses are a function of how the question is framed, whether in terms of lives saved or of lives lost.

Id.

44. *See id.* at 33.

45. Statistical probabilities enable physicians to evaluate new or comparative treatments, causes of disease, and potential diagnoses. *See* PETER SPRENT, TAKING RISKS: THE SCIENCE OF UNCERTAINTY 200 (1988).

46. Conversation with Lila Singer, in Cambridge, Mass. (Dec. 16, 1991); *see also* C.S. Peirce, *The Red and the Black*, *in* 9 GATEWAY TO THE GREAT BOOKS 342 (1963). Peirce states:

Given these factors, what kinds of medical conversations about statistical assessments can reflect ethical regard for patients? If a patient has undergone a screening test and produced a positive result, how should the physician communicate that result, given the chance that the positive result is false? At a minimum, the doctor needs to understand the probabilities or statistical pattern. Next, it would help if the doctor discerned ways to communicate the information that would match or refer to what the patient does know and understand. One possibility would be to look for images from the patient's life experience that can make the probabilities concrete.[47] The patient intuitively deals with and compares probabilities everyday, including the risk of car accidents on a busy road compared with the risks on a quiet street, or the chance of rain rather than snow in February.[48]

It would also help express the physician's concern for the patient to acknowledge the fear the statistical information potentially creates for the patient, and to discuss explicitly, as a task to be shared, the effort to make sense of uncertainty.[49] Where the patient needs to make a decision by comparing risks of alternate treatments, using narrative descriptions and crafting alternate arguments rather than merely repeating the quantitative risks could be helpful.[50] The doctor could offer in words the kinds

---

> An individual inference must be either true or false, and can show no effect of probability; and, therefore, in reference to a single case considered in itself, probability can have no meaning. Yet if a man had to choose between drawing a card from a pack containing twenty-five red cards and a black one, or from a pack containing twenty-five black cards and a red one, and if the drawing of a red card were destined to transport him to eternal felicity, and that of a black one to consign him to everlasting woe, it would be folly to deny that he ought to prefer the pack containing the larger portion of red cards, although, from the nature of the risk, it could not be repeated.

Id. at 345.

47. Conversation with Dr. Linda Emanuel, in Cambridge, Mass. (Feb. 20, 1992).

48. Some people might be assisted by visual depictions of probabilities or even of large numbers. *See* EDWARD R. TUFTE, ENVISIONING INFORMATION 29, 43 (1990); EDWARD R. TUFTE, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION 91 (1983).

49. *Cf.* SPRENT, *supra* note 45, at 254 ("The importance of psychological reactions to risk cannot be ignored, but the sociologist and psychologist have a duty to ensure that emotion is tempered by reason rather than fueled by illogical fear. This aim will be achieved if the layman can be given a better appreciation of uncertainty.").

50. For a related tactic, see HOFFMANN, *supra* note 2, at 130-131 ("The best solution I found was to avoid giving any direct advice but, instead, to offer people sample choices pro and con that other patients had made in similar situations in the past," including

of considerations other patients have considered and thereby connect the medical choice with an array of values and practical concerns the patient may have, while also connecting the patient with a sense of the community of patients who have faced similar decisions.

But underlying any particular recommendation is the ethical concern posed by communicating technical and statistical information: the physician must attend to the meanings and misunderstandings the patient may experience while trying to elicit informed consent. This involves getting to know enough about the patient's methods for processing information, and the patient's methods for making sense of the world, to translate the technical information in a way that can be grasped. If this involves crossing over from one culture, a technical, scientific one, to another, then that travel should be part of the ethical physician's conduct.[51]

C. Telling Too Much?

It is possible that physicians at times may tell too much information. In trying to fulfill the demands of informed consent or in trying to avoid any implication that the doctor knows best, doctors may share much or all of the information with a patient in the form that the doctor obtains it. Here arises a paradox in the relation between information and control. If once the physician's nondisclosure of information excluded the patient from control over medical judgments, full disclosure of information—complete with probabilities and uncertainties—may engender no increased sense of control for patients. Perhaps for this reason, doctors do edit and condense information when talking with patients.[52]

This practice can be justifiable when it reflects actual knowledge of the specific patients, taking into account the potential response of patients to worrisome or confusing informa-

---

reasons for and against a given treatment, "which were like courtroom arguments that [patient's] could accept or reject.").

51. *See supra* note 29 and accompanying text (discussing C.P. Snow's thesis that scientists and humanists inhabit separate worlds).

52. *See* HOFFMANN, *supra* note 2, at 131-132. Hoffmann notes: "Particularly vexing is the question of how far to go in discussing risks and benefits with patients." *Id.* at 131. Due to considerations of practicality and mercy, physicians usually abridge the laundry list of risks and benefits whenever discussing options with patients. *Id.* at 132.

tion. Simply telling all as a general practice does not fulfill the ethical commitment to the particular person who is the patient; different patients hear and understand differently. One physician addresses this point by considering with each patient the range of potential ways of characterizing a diagnosis.[53] He reports that he picks the version to tell the patient based on his own effort to "normalize" the patient's response; he seeks to produce a response with the proper level of concern and at the same time without too much anxiety. For example, a patient's presenting symptoms could be characterized alternatively as a bad cold, bronchitis, or walking pneumonia. This doctor explains that he selects which characterization to share with the patient using his own knowledge of patients; he would report walking pneumonia to an individual he knows is disinclined to take care of his health and a bad cold to one who is prone to anxiety.[54]

This is one way to incorporate actual knowledge of specific individuals when the physician discusses medical information with a patient. More serious and complex diagnoses and treatment choices would pose more difficult judgments about how to communicate. But the principle would be the same. The doctor should remember the person in the telling and recognize the consequences for that person's sense of control and dignity. The doctor needs to balance these concerns with attention to the opposite danger of misrepresentation through silence,[55] and with the risk of constricting patient control by restricting access to information.[56] It is these kinds of competing concerns that make ethics a continu-

53. Conversation with Dr. Robert Singer, in Cambridge, Mass. (Feb. 28, 1992).

54. Dr. Singer acknowledges that each of these judgments reflects his own notion of the right level of concern the patient should have about the condition, but this is part of the expertise and knowledge the patient seeks from the doctor. *Id.*

55. Annas, *supra* note 4, at 364. Annas asserts: "Physicians can mislead patients as much by silence as by direct advice. In the doctor-patient context, a half-truth is the same as a lie, and it violates both medical ethics and the doctrine of informed consent." Annas applies this point in criticizing the Supreme Court's decision in Rust v. Sullivan, 111 S. Ct. 1759 (1991), which approved regulations forbidding physicians in clinics receiving federal funds from discussing abortion: "By legally approving inherently unethical behavior, the Court's opinion in *Rust* is a direct attack on medical ethics in the doctor-patient relationship." *Id.*

56. *See* TODD, *supra* note 18, at 96 (discussing danger when doctors do not give enough technical information and women are thus unable to make adequately informed decisions about birth control).

ing, complex process of analysis and response rather than a set of preannounced rules.[57]

II. TELLING FAMILIES

If the ethical dimensions of doctor-patient communications were not difficult enough, the complexity deepens with the presence of patients' families. How should physicians respond when a patient does not want the family to know about the medical diagnosis or treatment choices? How should physicians respond when the family does not want the patient to know? And how should families act when the patient cannot be told or cannot decide due to the severity of the illness or another incapacitating condition? All of the difficulties with telling not enough, telling too technically, and telling too much are compounded by the involvement of family members. Complexities accompanying the report of statistical and other technical information are certainly compounded when more people, with their own ways of understanding and misunderstanding, are involved. Let me explore several specific difficulties presented by the involvement of family members.

First, the patient may not want the family to know about a serious diagnosis. May Sarton's novel, *A Reckoning*, explores this possibility through a middle-aged character who fears a loss of control and dignity over her own death if her children are informed of her illness.[58] The novel affords a vivid glimpse of one person's effort to use knowledge of her terminal illness as an opportunity for personal and spiritual growth while suggesting that, at least in this case, intervention by her family would have curtailed that opportunity.

Similarly, a woman who discovers she is pregnant may well wish to keep this information from her family as she decides whether to terminate the pregnancy. That anti-abortion activists have targeted this very moment as a point of state intervention reveals the close connection between control and information; mandated disclosures to the parents of a minor who is pregnant, or to the spouse or male partner of a pregnant adult, alter the

57. *Cf.* David Wilkins, *Who Should Regulate Lawyers?*, 105 HARV. L. REV. 801 (1992) (discussing professional responsibilities of lawyers).

58. MAY SARTON, A RECKONING (1978).

patient's own relationship to the pregnancy and often her decision making about it.[59] If such laws were to direct the physician to notify family members, they would interfere with the physician's ethical commitment to the patient. For that commitment, guided by the norm of informed consent and the Golden Rule, directs the physician to respect the patient's autonomy, dignity, and freedom from harm.

I will leave for another day arguments for keeping the state out of such matters. Assuming for the moment that there is no legally mandated disclosure to family members, what should a doctor do when a patient wishes not to disclose her diagnosis or treatment options to her family? In the situation depicted by May Sarton, the doctor urges the patient to inform her adult children, due to a sense that they should know. When pressed by the patient not to pursue this point, the doctor realizes that more important than telling the children is ensuring that the patient makes arrangements for her own care in advance of the time that she loses her self-sufficiency. This realization by the doctor expresses the kind of insight I would urge more generally. The physician can raise the matter of sharing information with family members, but should listen to the patient's response, and bring the patient to acknowledge underlying needs rather than insist on a particular way to meet them.

The second difficulty arises when a family member learns of the patient's condition first and then does not want the patient to know. Physicians may disclose the results of surgery or screening tests to spouses, parents, adult children, or others deemed by the patient to be family members.[60] When the diagnosis is, for example, Alzheimer's disease, family members may urge the physician not to tell the patient on the grounds that the sheer knowledge of this progressive, debilitating disease would produce pain and indignity, and the patient ultimately will lose the ability to understand anyway.

A countervailing argument, however, stems from the betrayal of trust the patient could experience if the doctor abides by the

---

59. *See* H.L. v. Matheson, 450 U.S. 398, 436-41 (1981) (Marshall, J., dissenting).

60. A critical question for medical practice in these contexts is exactly who is the family; will the doctor recognize a cohabitant, a close friend, a same-sex lover? For arguments in favor of deference to the patient's definition of family, see Martha Minow, *In All Families: Owing and Loving*, 95 W. VA. L. REV. (forthcoming Winter 1993).

family's wish to shield the patient from the truth. One medical researcher attributes his life's work to the crisis he experienced as a teenager when he watched his grandfather's dying: "As he was dying, his doctors maintained a circle of lies and distortions that left him confused and dispirited. Other well-meaning members of my family cooperated in these communicative maneuvers, all intended to keep up his morale and our family's ability to function."[61]

Based on similar circumstances, Dr. Jay Katz concludes that "withholding information about the gravity of patients' conditions can only confuse patients, since the rosy picture the doctor is projecting does not agree with their own intuitive evaluation of their condition. It can leave patients with a nagging sense of doubting their own judgment, if not their sanity."[62]

Dr. Hoffmann recounted this contrasting example:

> [T]he patient was an elderly Hispanic gentleman with stomach cancer who had been referred to me for diabetes and high blood pressure. Although his family had learned of his diagnosis when the cancer was first noted incidentally during an unrelated operation, the patient had never been told of the condition. When I assumed responsibility for his medical care, his relatives asked me to join in a long conspiracy of medical silence because of the implications that cancer had within his culture. To have cancer, they explained, meant that one had not lived well, and they argued that making the diagnosis known to him would only cause him added and unnecessary suffering.[63]

---

61. WAITZKIN, *supra* note 19, at xiii.

62. KATZ, *supra* note 1, at 224-25. He continues:

> One dying woman expressed this well, in demeanor and words, when her doctor finally told her that she was not "nervous" but dying: "The smile she gave [her doctor] actually expressed relief. 'Thank God,' she said, 'Someone's finally told me the truth.'" The truth about her impending death, however painful, liberated her from isolation; from the madness of doubting her own sanity; from the abandonment that she experienced as she listened to her doctors, nurses, and family, all of whom talked largely about trivialities and not about real concerns; and from the abandonment created by furtive contacts to avoid revealing the truth inadvertently.

Id. at 225 (quoting Thomas P. Hackett & Avery D. Weisman, *When to Tell Dying Patients the Truth*, MED. ECON., Dec. 4, 1961, at 81).

63. HOFFMANN, *supra* note 2, at 139.

Hoffmann did not report his response.[64]

The position taken by Dr. Katz, and implied by the informed consent doctrine, would direct the physician in these circumstances to tell the patient. Only this route would respect the patient's autonomy and right to self-determination. Yet, the analysis I have been pursuing here advises the doctor to learn about the particular patient in framing an ethical response. This might well point to deference to the family's judgment. Yet that conclusion is rendered less than secure by the fact that it is the family rather than the patient who provided the information affecting the doctor's judgment about whether to tell the individual medical information. The better practice, I think, would engage the doctor in an effort to get to know the patient as an individual, and to determine whether he wants to know information of the sort represented by the cancer diagnosis. Perhaps, just as advance directives allow patients to indicate their treatment preferences before the issue arises and before any loss of their own competence,[65] doctors could discuss with competent patients their own desires about being told medical information, and about letting their families know such information. Building into the doctor-patient relationship the conversation about how much the patient wants to know and wants family members to know would help construct the ethical practice I urge here.

Perhaps most well-known to lawyers is the third permutation of family involvement with medical care: when the patient cannot speak, how can or should the family speak? This question has animated landmark judicial opinions in biomedical ethics,[66] and much scholarly writing.[67] Again, I do not mean here to engage in

---

64. *Id.*

65. *See generally* ALLEN E. BUCHANAN & DAN W. BROCK, DECIDING FOR OTHERS: THE ETHICS OF SURROGATE DECISION MAKING 95-112, 152-159 (1989) (discussing ethical ramifications of advance directives).

66. *See, e.g.*, Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2845 (1990) (rejecting parent's petition to terminate persistent vegetative daughter's artificial nutrition and hydration); Brophy v. New England Sinai Hosp., 497 N.E.2d 626, 639-40 (Mass. 1986) (honoring wife's decision to discontinue persistent vegetative husband's artificial nutrition and hydration); *In re* Quinlan, 355 A.2d 647, 671-72 (N.J.), *cert. denied*, 429 U.S. 922 (1976) (granting father power to authorize discontinuance of persistent vegetative daughter's life-support apparatus).

67. *See, e.g.*, BUCHANAN & BROCK, *supra* note 65 (discussing ethical issues in surrogate decision making); DONALD VANDEVEER, PATERNALISTIC INTERVENTION: THE MORAL BOUNDS ON BENEVOLENCE 345-421 (1986) (discussing family intervention and

debates over the legal standards applicable to such circumstances. Rather, to enhance attention to the ethical concerns by both the medical personnel and the family, I have a modest suggestion for those patients who once had awareness and competence.[68] It would have to be accomplished well in advance of medical crises. Family members or friends should consider producing video or audio interviews with the individual so that these materials are available for those who may later be involved in the individual's medical treatment.[69] Without pointing in one direction or another regarding choices of treatment, this may be one way to remind those involved of the particular human being who is now the patient.

III. THE DIFFERENCE BETWEEN GETTING KNOWLEDGE AND BEING KNOWN

Anatole Broyard, a columnist, described his ideal doctor as someone compassionate, someone who could almost serve as confessor, someone who could let the patient feel known as a person, not only as a patient.[70] Getting knowledge is different from being known; if Broyard speaks for others besides himself, physicians should listen to patients' needs not just for knowledge but also for acknowledgment as individuals.

Being known, though, requires more than getting evidence that another knows something about you. This reminds me of the story of the wealthy man who had made gifts to a hospital at which he had been a patient years before.[71] Then, subsequently,

---

informed consent); Elizabeth S. Scott, *Sterilization of Mentally Retarded Persons: Reproductive Rights and Family Privacy*, 1986 DUKE L.J. 806 (discussing reproductive rights of incompetent people).

68. Advance directives and proxies drawn up by such individuals would satisfy the legal requirements for those communities that require clear evidence of the no-longer-competent patient's wishes about treatment.

69. *Cf.* Ellen J. Miller, *Any Family Can Make History*, NEWSWEEK, Dec. 23, 1991, at 9 (describing effect of videotape made of mother on nurses who cared for her at the end of her life).

70. ANATOLE BROYARD, INTOXICATED BY MY ILLNESS 42-44 (1992); *see also* PETER FEIBLEMAN, LILLY: REMINISCENCES OF LILLIAN HELLMAN 292 (1988) (discussing insensitivity of distinguished specialists caring for Lillian Hellman); ARTHUR W. FRANK, AT THE WILL OF THE BODY: REFLECTIONS ON ILLNESS 10-12 (1991) (discussing limitations of physicians' professional personae in addressing needs of patients undergoing life-altering experiences).

71. GOROVITZ, *supra* note 38, at 28-29.

he arrived at the emergency room in need of an X ray for a twisted ankle. As a former patient, he presented to the receptionist a patient identification card; the receptionist typed his number into her computer terminal and this conversation ensued:

"'Mr. S., do you still live at the same address on William St.?'"

"'Yes, I do.'"

"'And you're going to give your organs to the hospital?'"

"Taken aback, he replied, 'Well, yes. But not today. It's only an injured ankle! What prompted you to ask me that?'"

"'Oh. Well, it says here on the screen that you're a major donor.'"[72]

Even beyond incidents of this nature, some kinds of knowledge obtained through medical encounters is inappropriately shared. More important to my inquiry, though, than efforts to control access to information, is the stance of the individual who faces a medical encounter. Can the doctor overcome the risk of indifference that comes with the passage of time, the burdens of case loads, and the self-protection needed to do the job? It might help if doctors worked on viewing patients as the "principal characters in their own living dramas" rather than the site of a disease or the subject of statistical data.[73] This is not an argument against technical knowledge, just a reminder that ethics requires other kinds of knowledge.[74]

The ethical commitment must work in both directions. Patients and family members, as well, need to view doctors as human and as individuals rather than god-like creatures who are

---

72. *Id.*

73. The phrase comes from HOFFMANN, *supra* note 2, at 15, and grew out of his course in medicine and literature taught by Robert Coles. *See generally* ROBERT COLES, THE CALL OF STORIES: TEACHING AND THE MORAL IMAGINATION (1989) (discussing the use of stories in teaching and moral development). Yet Hoffmann may take the idea too far when he writes that he has become a literary critic, alert to symbolism and allusion, hidden meanings in people's stories, and attentive to different schools of thought in literary interpretation, *id.* at 170, for this very intellectual approach may produce the kind of remoteness the analogy to literature was supposed to remedy.

74. *See* PAULOS, *supra* note 41, at 90-92 (arguing it is wrong to think that numbers diminish human individuality, or undermine sense of big questions). For an insightful effort to advance medical ethics, see Ezekiel J. Emanuel & Linda L. Emanuel, *Four Models of the Physician-Patient Relationship*, 267 JAMA 2221 (1992) (comparing paternalistic, informative, interpretative, and deliberative models of physician-patient relationships and rejecting the transmission of information as physician's sole goal because it reduces the physician to a technologist).

all-knowing—and rather than villains or bearers of bad news.[75] Doctors have their own experiences of feeling overwhelmed, uncertain, and powerless. Sharing these feelings immediately with patients and family members may not be wise,[76] but the more genuine the communication in all directions, the greater chance for some shared understanding and appreciation. Telling stories to one another may be an important path toward converting patterns of authority and manipulation into opportunities to be known.

75. In this way, patients and family members can help doctors accomplish the transformation urged by Dr. Jay Katz. But Dr. Katz warns: "Conversation, however, will protect the integrity of the physician-patient relationship only if doctors are willing to confront and change their views of themselves as sole authority and of their patients as incompetent participants in decision making. Otherwise, manipulation and coercion will continue to rule their interactions." KATZ, *supra* note 1, at 164.

76. However, greater sharing among doctors might help overcome not only feelings of isolation, but also feelings of discomfort with the uncertainty that pervades medicine. *See* GOROVITZ, *supra* note 38, at 173 (noting that due to lack of time and encouragement to reflect openly about uncertainties and anxieties, senior and junior medical colleagues experience the same kind of silence that Dr. Katz found between doctors and patients).

Utah's Statute of Limitation Barring Minors from Bringing Medical Malpractice Actions: Riding Roughshod over the Rights of Minors?

> One of the basic tenets of our system of law and justice is that it attempts to accord to all individuals protection in their persons and property, and this is true, a fortiori, of children.[1]

I. INTRODUCTION

Since before statehood, Utah has placed children in a distinct position of legal disability,[2] prohibiting them from personally bringing a cause of action before reaching the age of majority.[3] The purpose of according children this status as "minors"[4] is not to deprive them of rights, but to safeguard their rights during a period in which they are viewed as immature, inexperienced, and unable to protect their rights independently.[5] Utah has always partially compensated for this disability by allowing a parent or guardian to act on behalf of a minor to bring a minor's claim in court.[6] To safeguard the rights of minors in cases where a parent or guardian fails to act on behalf of the minor, Utah has traditionally exempted minors from statutes of limitation during their period of legal disability.[7] As a result, Utah has almost always

---

1. State *ex rel.* K.B., 326 P.2d 395, 396 (Utah 1958).

2. *See* Act of Feb. 6, 1852, ch. 3, § 1, 1876 UTAH COMP. LAWS 345, 345 (current version at UTAH CODE ANN. § 15-2-1 (1986)).

3. *See* Civil Practice Act, tit. 1, § 9, 1876 UTAH COMP. LAWS 362, 368 (current version at UTAH R. CIV. P. 17(b)).

4. For purposes of this Comment, the term "minor" refers to a person below the legal age of majority. *See, e.g.*, UTAH CODE ANN. § 15-2-1 (1986) (providing that period of minority for Utah children extends until age 18, but all minors obtain majority by marriage).

5. *See* Robert M. Horowitz & Beth G. Hunter, *The Child Litigant*, *in* LEGAL RIGHTS OF CHILDREN § 3.02, at 73-74 (Robert M. Horowitz & Howard A. Davidson eds., 1984); 42 AM. JUR. 2D *Infants* § 1 (1969).

6. *See* Civil Practice Act, tit. 1, § 9, 1876 UTAH COMP. LAWS 401, 402 (current version at UTAH R. CIV. P. 17(b)) (minor's actions must be brought by a guardian).

7. *See* Act of Feb. 16, 1872, ch. 4, § 24, 1876 UTAH COMP. LAWS 362, 368 (current version at UTAH CODE ANN. § 78-12-36 (1987)) (tolling the statute of limitations for minors); *cf.* Horowitz & Hunter, *supra* note 5, § 3.04 (discussing tolling of statutes of limitation for minors).

accorded minors the right to bring an action on their own behalf upon reaching the age of majority.[8]

In response to a perceived crisis in medical malpractice liability insurance, however, the 1979 Utah Legislature expressly eliminated the tolling provision for minors' legal disability in the context of medical malpractice claims.[9] The Legislature sought to limit the adverse effects of rising malpractice liability insurance premiums, in part, by holding minors to the same medical malpractice limitation period as adults.[10] In doing so, the Legislature completely abrogated minors' rights to personally bring medical malpractice cases, requiring them to rely solely on parents or guardians to assert their rights for them.

This Comment critiques the Legislature's action prohibiting minors from bringing medical malpractice claims and argues that the Legislature's position violates both the open courts and equal protection guarantees provided by the Utah Constitution. This Comment recognizes the serious, problematic issues of access to medical care the current law attempts to address. Nonetheless, the current law's restriction on minors is arguably unconstitutional because it substantially infringes on minors' rights, while providing only questionable benefits to society.

To reach this conclusion, this Comment first examines the historical background of limitation statutes, disability tolling provisions, and minors' rights to personally bring causes of action.

A statute of limitation establishes a specific period of time during which a claim must be brought. Raithaus v. Saab-Scandia of Am., Inc., 784 P.2d 1158, 1160 (Utah 1989). This Comment distinguishes a statute of limitation from a statute of repose, which also places a time limit on a plaintiff's cause of action. A statute of limitation allows a plaintiff to file a lawsuit only within a specified period of time after a legal right has been violated. *Raithaus*, 784 P.2d at 1160. For example, in the area of medical malpractice, a statute of limitation would begin running at the time an injured plaintiff knows or should have known of the injury. *See* Foil v. Ballinger, 601 P.2d 144, 147 (Utah 1979). A statute of repose, on the other hand, is "designed to bar actions after a specified period of time has run from the occurrence of some event other than the injury which gave rise to the claim." *Raithaus*, 784 P.2d at 1160. For example, a statute of repose would begin to run at the time of the medical treatment regardless of when the plaintiff discovers an injury resulting from that treatment. *See id.*

8. *See infra* note 31 (discussing the few historical exceptions to tolling for minors under Utah law).

9. *See infra* notes 34-51 and accompanying text (discussing passage of Utah Health Care Malpractice Act).

10. *See infra* notes 34-51 and accompanying text (discussing passage of Utah Health Care Malpractice Act).

It next reviews recent Utah cases scrutinizing the constitutionality of legislation challenged under Utah's open courts and equal protection provisions. It applies the tests outlined in these recent cases to the legislation eliminating the tolling provision for minors in medical malpractice actions, and illustrates that the Utah Supreme Court would likely find such treatment unconstitutional. Finally, this Comment concludes with a brief review of alternative legislative approaches that strike a more equitable balance between minors' rights and the need to address malpractice insurance concerns.

II. THE RIGHTS OF MINORS VERSUS THE NEED FOR STATUTES OF LIMITATION

A. A General View Favoring Minors' Rights

For centuries, statutes of limitation have served practical public policies by barring litigation of stale and fraudulent claims, thereby improving judicial administration and fairness to defendants.[11] Apart from enhancing efficient judicial administration, statutes of limitation have been used for policy reasons to disfavor particular types of actions.[12] The obvious consequence of these

---

11. Wood v. Carpenter, 101 U.S. 135, 139 (1879); *see also* Horton v. Goldminer's Daughter, 785 P.2d 1087, 1091 (Utah 1989) (describing public policies underlying statutes of limitation). The United States Supreme Court has stated:

> [s]tatutes of limitations are primarily designed to assure fairness to defendants. Such statutes "promote justice by preventing surprises through the revival of claims that have been allowed to slumber until evidence has been lost, memories have faded, and witnesses have disappeared. The theory is that even if one has a just claim it is unjust not to put the adversary on notice to defend within the period of limitation and that the right to be free of stale claims in time comes to prevail over the right to prosecute them." Moreover, the courts ought to be relieved of the burden of trying stale claims when a plaintiff has slept on his rights.

Burnett v. New York Cent. R.R., 380 U.S. 424, 428 (1965) (quoting Order of R.R. Telegraphers v. Railway Express Agency, 321 U.S. 342, 348-49 (1944)).

Statutes have played this role in common law actions for at least 350 years. *See Developments in the Law—Statutes of Limitations*, 63 HARV. L. REV. 1177, 1178 (1950); *see also* Housing Auth. of Union City v. Commonwealth Trust Co., 136 A.2d 401, 404 (N.J. 1957) (first statutes of limitation passed by English Parliament when abuses from stale demands became burdensome). Statutes of limitation found in modern American law apparently derive from the English statute of 21 Jam. 1, ch. 16 (1623). *See Wood*, 101 U.S. at 139; JOSEPH K. ANGELL, LIMITATIONS OF ACTIONS AT LAW 10 (5th ed., Boston, Little, Brown & Co. 1869).

12. *Developments in the Law—Statutes of Limitations*, *supra* note 11, at 1180. The

statutes is the occasional harsh result of barring otherwise legitimate claims.[13] A legislature's power to establish or alter a statute of limitation is well recognized, however, and courts generally respect a legislature's decision on such a policy matter.[14] Nevertheless, statutes of limitation are clearly subject to constitutional constraints.[15]

Policies supporting statutes of limitation conflict with the general policy of according minors a special status designed to protect their legal rights. While children may sue and be sued,[16] minors are often deemed "legally disabled," which prevents them from personally bringing an action in court.[17] The purpose of this legal disability is not to defeat minors' rights, but to protect their rights during a period characterized by immaturity and inexperience.[18] This well-established policy of protecting the rights of

---

statute of limitation that is the topic of this Comment is an example of a statute of limitation being used to disfavor a type of action—medical malpractice—for policy goals other than judicial convenience or avoiding the evidentiary problems of stale claims. *See infra* note 50 (purpose in restricting statute of limitation period for medical malpractice claims is to minimize adverse effects of medical malpractice litigation on access to public health care).

13. *See* Christiansen v. Rees, 436 P.2d 435, 437 (Utah 1968) (Henriod, J., dissenting) ("It is obvious that limitations statutes, in their role of repose, may lead to harsh results."); *see also* Chase Securities Corp. v. Donaldson, 325 U.S. 304, 314 (1945) (statutes of limitation are by definition arbitrary and operation does not discriminate between just and unjust claims).

14. *See Chase*, 325 U.S. at 314; Saranac Land & Timber Co. v. Roberts, 177 U.S. 318, 323-24 (1900); Toronto v. Sheffield, 222 P.2d 594, 596 (Utah 1950).

15. *See, e.g.*, Pickett v. Brown, 462 U.S. 1, 18 (1983) (invalidating statute of limitation as unconstitutional); *cf.* Horton v. Goldminer's Daughter, 785 P.2d 1087, 1094 (Utah 1989) (invalidating statute of repose as unconstitutional).

16. Horowitz & Hunter, *supra* note 5, § 3.01; *see, e.g.*, Blum v. Stone, 752 P.2d 898, 899-900 (Utah 1988) (allowing a guardian to bring minor's medical malpractice claim).

17. *See* Horowitz & Hunter, *supra* note 5, § 3.02; *see, e.g.*, UTAH R. CIV. P. 17(b) (barring minors from bringing actions on own behalf). Unlike Utah, most states and the federal rules do not require minors to have a representative, but they are typically provided one in the court's discretion. *See* Horowitz & Hunter, *supra* note 5, § 3.03; *see, e.g.*, Fed. R. Civ. P. 17(c) (court shall appoint a guardian ad litem or such other orders as the court deems proper for the infant's protection). A court may decline to appoint a representative, however, when the minor is an older child capable of understanding and participating in the action, and is represented by counsel. Horowitz & Hunter, *supra*, § 3.03.

18. Horowitz & Hunter, *supra* note 5, § 3.02. According to these authors,

> [d]ue to their age, immaturity, and inexperience, children are disadvantaged in conducting adult business. There are few instances in which this is more true than in the business of litigation, an undertaking that is baffling to most adults. By definition, therefore, children have been deemed legally incompetent or disabled with respect to legal procedures .

children comes into conflict with policies supporting statutes of limitation in cases where such statutes bar minors' claims before minors have the legal capacity to bring them.[19] This conflict is partially resolved by allowing minors to bring claims through an adult representative.[20] As a result, minors with someone to represent them do not have to wait until reaching the age of majority to seek a remedy in court.

Providing adult representation alone, however, is potentially inadequate to protect the rights of minors because parents or guardians may fail to bring a minor's cause of action within the required limitation period.[21] Accordingly, as a second method for assuring minors' access to courts, legislatures generally extend statute of limitation periods to allow minors to enforce their rights upon reaching the age of majority.[22] By using a tolling statute, the law ensures "that a minor's rights are not lost because a parent or guardian neglected to protect the minor by bringing a timely action."[23]

. . .

Id.; *see also* Doe v. Durtschi, 716 P.2d 1238, 1248 (Idaho 1986) ("Minors lack the judgment, experience, and awareness to protect their rights with appropriate, timely civil action . . . ."); Boyle v. Kirby (*In re* Davidson's Will), 26 N.W.2d 223, 225 (Minn. 1947) (disabilities used not to defeat minors' rights "but to shield and protect them from the acts of their own improvidence as well as from acts of others"); Hunter v. North Mason High Sch., 529 P.2d 898, 899 (Wash. Ct. App. 1974) (legal disabilities of minors established for their protection, and not as a bar to enforcement of their rights).

19. *See generally* Haakanson v. Wakefield Seafoods, Inc., 600 P.2d 1087, 1090 (Alaska 1979) (discussing legislature's weighing protection of minors' rights against policies underlying statutes of limitation).

20. *See* Horowitz & Hunter, *supra* note 5, § 3.02; *see, e.g.*, UTAH R. CIV. P. 17(b) (minor's cause of action may only be brought by general guardian, or guardian ad litem appointed in the particular case by the court).

21. *See* Dye v. Fremont County Sch. Dist., 820 P.2d 982, 985 (Wyo. 1991) ("In the interest of justice, we cannot allow a minor, who has no realistic ability to protect herself, to suffer loss of her claim because of a parent's failure to act."); *see also Durtschi*, 716 P.2d at 1248 ("Minors should not have to rely upon others to protect their rights."); Wilbon v. D.F. Bast Co., 382 N.E.2d 784, 790 (Ill. 1978) (holding child incapable of initiating enforcement proceeding will not be solely dependent on "whim or mercy" of representative to protect child's rights); Lazich v. Bellanger, 105 P.2d 738, 739 (Mont. 1940) (minors right of action should not be frittered away because of parent's omission); *Hunter*, 529 P.2d at 900 ("[Minor's] right of action should not depend on the good fortune of having an astute relative or friend to take the proper steps on his behalf.").

22. *See Durtschi*, 716 P.2d at 1250 ("The obvious intent of [the general tolling provision] is to preserve the rights of injured minors until they are old enough to take appropriate action."); *see, e.g.*, UTAH CODE ANN. § 78-12-36 (1992) (allowing minors to toll limitation statutes during time of minority).

23. Scott v. First State Ins. Co., 456 N.W.2d 152, 155 (Wis. 1990) ("The purpose of

Apparently, lawmakers have long believed that minors' legal rights should not be destroyed by the failure of a parent or guardian to bring a child's cause of action, since for over 360 years statutes of limitation have generally included tolling provisions for minors.[24] This view is still widely held today. The great majority of states have general tolling statutes that suspend the operation of almost all limitation periods for minors until they reach the age of majority.[25] Overall, therefore, it appears that state legislatures have resolved the conflict between the policies underlying statutes of limitation and the protection of children's rights largely in favor

---

tolling the statute of limitation when the plaintiff is a minor is to ensure that a minor's rights are not lost because a parent or guardian neglected to protect the minor by bringing a timely action."); *see also* Cross v. Pacific Gas & Elec., 32 Cal. Rptr. 504, 509 (Ct. App. 1963), *vacated on other grounds*, 388 P.2d 353 (Cal. 1964) (tolling provision intended to protect minors' rights and such rights should not be lost because parents slept on rights); Thea Andrews, *Infant Tolling Statutes in Medical Malpractice Cases: State Constitutional Challenges*, 5 J. LEGAL MED. 469, 469 (1984) (rationale for tolling statutes is that minors should not be penalized if parent or guardian fails to bring minor's claim); *see generally* Horowitz & Hunter, *supra* note 5, § 3.02 (recognizing tolling statutes as "very important" procedure assuring children's access to courts).

Judicial doctrines similarly reflect the concern that a parent or guardian may not adequately represent a minor in court. For example, whenever a minor appears before a court, it is the court's established duty to independently assure that the minor's interests are properly represented and protected. *See, e.g.*, Stone v. Gulf Am. Fire & Casualty Co., 554 So. 2d 346, 361 (Ala. 1989) (court has duty to protect rights and interests of ward of court even though ward has appointed representative); Schierenbeck v. Minor, 367 P.2d 333, 334 (Colo. 1961) (state policy to insure minors' rights are not impaired by failure of courts and attorneys to sufficiently present their cause); Hun v. Center Properties, 626 P.2d 182, 189 (Haw. 1981) ("[I]t is the role of the courts to protect the interests of minors who become parties to litigation."); Severs v. Country Mut. Ins. Co., 434 N.E.2d 290, 292 (Ill. 1982) (minor litigants entitled to special protection by courts, especially to ensure that their rights are protected from negligence of their representatives). Also, a majority of jurisdictions toll the limitation period for a minor if the minor's guardian brings the minor's claim but subsequently fails to pursue it to completion. 54 C.J.S. *Limitations of Action* § 114 (1987); 51 AM. JUR. 2D *Limitations of Actions* § 183 (1970); *see generally* Jean E. Maess, Annotation, *Tolling of State Statute of Limitations in Favor of One Commencing Action Despite Existing Disability*, 30 A.L.R. 4TH 1092, 1093 (1984) (describing general rule). Such principles evidence a judicial determination (as opposed to a legislative position) that it may be inadequate to rely solely on parents or guardians to protect minors' rights in court.

24. *Developments in the Law—Statutes of Limitations*, *supra* note 11, at 1229. The English statute 21 Jam. 1, ch. 16 (1623) is credited as the forerunner to modern statutes of limitation. The statute included liberal statutory exceptions for legal disabilities, including infancy. *See* ANGELL, supra note 11, at 194; *Developments in the Law—Statutes of Limitations*, *supra* note 11, at 1178.

25. Horowitz & Hunter, *supra* note 5, § 3.04; *see also* 1 DAVID W. LOUISELL & HAROLD WILLIAMS, MEDICAL MALPRACTICE ¶ 13.13 n.40 (1991) (general tolling provision found in 48 states). Other legal disabilities include mental incompetence and incarceration. *Id.*

of children's rights. Consequently, statutes that toll limitation periods represent the generally accepted principle that the legal rights of children should not be destroyed by a statute of limitations.[26]

This general view favoring protection of minors' access to courts is not without significant exception. Under federal law, for example, there is no constitutional barrier to placing children on the same footing as adults for purposes of statutes of limitation.[27] As a result, lawmakers occasionally reject the general view favoring minors' rights and deny minors the protection of tolled limitation periods.[28] Nevertheless, given the long-standing and

---

26. *See, e.g.*, Haakanson v. Wakefield Seafoods, Inc., 600 P.2d 1087, 1090 (Alaska 1979) (tolling provisions express public policy of safeguarding minors' rights); Williams v. Los Angeles Metro. Transit Auth., 440 P.2d 497, 499 (Cal. 1968) (tolling statutes express deep and long-recognized principle that children are protected during minority from destruction of rights by running of statutes of limitation); Doe v. Durtschi, 716 P.2d 1238, 1250 (Idaho 1986) ("[T]he obvious intent of the general tolling provision is to preserve the rights of injured minors until they are old enough to take appropriate action."); Thornton v. Mono Mfg. Co., 425 N.E.2d 522, 527 (Ill. App. Ct. 1981) (rationale underlying tolling statute for minors is that rights should not be extinguished merely because minors are not old enough to protect themselves); Roe v. Doe, 287 N.Y.S.2d 292, 298 (Fam. Ct. 1968) (infant should not be penalized by omission on part of guardian to sue on claim); *cf.* Scott v. School Bd., 568 P.2d 746, 747-48 (Utah 1977) (declaring limitation periods should be tolled for minors in all cases).

27. *See infra* notes 59 and 65 and accompanying text (discussing federal and state decisions finding constitutional certain medical malpractice exceptions to disability tolling provisions).

28. *See infra* note 59 (discussing federal exception to disability tolling provisions in Federal Tort Claims Act). As a general rule, courts will not provide a tolling exception for minors unless expressly provided by statute. *See, e.g.*, Johns v. Wynnewood Sch. Bd. of Educ., 656 P.2d 248, 249 (Okla. 1982) (unless excepted by statute, limitation period applies to minors); *see generally* 54 C.J.S. *Limitations of Actions* § 112 (1987) (tolling depends on statute and, unless excepted, minors treated same as adults); 51 AM. JUR. 2D *Limitations of Actions* § 182 (1970) (limitation period generally applies unless contrary to statute or provision). Once a general statutory tolling provision for minors exists, however, courts vary in construing how broadly that protection will apply. For example, some courts construe statutes granting a cause of action and containing their own limitation period to exclude a tolling provision for minors unless minors are expressly exempted from the particular limitation period. *See, e.g.*, Buder v. Merrill Lynch, Pierce, Fenner & Smith, Inc., 486 F. Supp. 56, 59 (E.D. Mo. 1980) (unless expressly mentioned by statute, minors are not exempted from limitation period in statute giving right of action and specific limitation period). Other courts, in attempting to reach a fair and reasonable construction of these statutes, have taken a more liberal view and examined relevant policy considerations. *See, e.g.*, Korth v. American Family Ins. Co., 340 N.W.2d 494, 497 (Wis. 1983) (courts interpreting statutes of limitation should be sensitive to protecting access to courts and other policy considerations); Barnett v. Clark, 448 N.E.2d 254, 255 (Ill. Ct. App. 1983) (courts should evaluate legislative intent and relevant policy considerations in determining whether statutorily-

widely-observed policy favoring minors' access to courts over the policies underlying statutes of limitation,[29] any attempt to extinguish the tolling provisions for minors should be carefully scrutinized.

B. Utah's Treatment of Minors Versus Statutes of Limitation

Historically, Utah law conforms with the general view favoring the protection of children's rights over the benefits of statutes of limitation. As early as 1876, Utah gave minors disability status and, accordingly, tolled statutes of limitation for minors.[30] Furthermore, Utah has rarely recognized exceptions to this protection.[31] Indeed, the Utah Supreme Court recently noted that "a minor claimant is justly entitled to the protection afforded by [the disability tolling statute], in all cases."[32] Despite an historical stance that was vigilantly protective of minors' rights, the 1979 Utah Legislature expressly eliminated the tolling provision for minors bringing medical malpractice claims.[33]

---

created cause of action is tolled for minors). *See generally* 54 C.J.S. *Limitations of Actions* § 6(b) (1987) (noting statutes of limitation interpreted liberally or strictly depending on jurisdiction).

29. *See, e.g.*, Steketee v. Lintz, Williams & Rothberg, 694 P.2d 1153, 1158 (Cal. 1985) ("Our courts have repeatedly recognized the strong public policy protecting minors against the loss of their rights due to the operation of statutes of limitation.").

30. *See supra* notes 2-8 and accompanying text (describing early Utah law regarding disability status of minors).

31. At one time, minors were required to comply with statutes allowing only a brief time period for giving notice of claims involving state or local government entities. *See* Scott v. School Bd., 568 P.2d 746, 747 (Utah 1977); *see also infra* notes 71-81 and accompanying text (discussing *Scott* holding). The Utah Supreme Court originally construed these statutes, using standard statutory interpretation principles, to apply to minors because they contained a specific time limitations period that took precedence over general provisions. *See* Varoz v. Sevey, 506 P.2d 435, 435-36 (Utah 1973); Gallegos v. Midvale City, 492 P.2d 1335, 1336-38 (Utah 1972). Within five years, however, the court reversed its position and held that the general tolling statute excepted minors from the notice requirement during the period of disability. *See Scott*, 568 P.2d at 748. Other than the Malpractice Act, research disclosed no other instance where the Utah Legislature has attempted to eliminate the tolling provision.

32. *Scott*, 568 P.2d at 748 (citation omitted).

33. UTAH CODE ANN. § 78-14-4(2) (1992).

III. THE MEDICAL MALPRACTICE INSURANCE CRISIS AND THE STATUTE OF LIMITATION RESPONSE

Medical malpractice insurance costs rose dramatically across the nation during the early 1970s.[34] As insurers abandoned a reputedly unprofitable and risk-laden medical liability market, concern arose over the continued availability of malpractice insurance.[35] Consequently, doctors and insurers lobbied state legislatures for measures that would restrict liability and reduce insurance premiums.[36] These groups argued that the impact of rising insurance costs substantially affected the cost and availability of health care.[37] They further argued that a substantial, if not primary, cause of the inflating cost of insurance was due to increases in the number and amount of malpractice claims and awards.[38]

---

34. Martin H. Redish, *Legislative Response to the Medical Malpractice Insurance Crisis: Constitutional Implications*, 55 TEX. L. REV. 759, 759-60 (1977); *see also* UTAH CODE ANN. § 78-14-2 (1992) (legislature finds dramatic increase in number of suits, claims for damages, and amount of judgments and settlements arising from health care).

35. Redish, *supra* note 34, at 760. *But cf.* U.S. DEPT. OF HEALTH, EDUCATION & WELFARE, PUB. NO. (OS) 73-88, MEDICAL MALPRACTICE: REPORT OF THE SECRETARY'S COMMISSION ON MEDICAL MALPRACTICE 38-40 (1973) [hereinafter HEW REPORT] (concluding malpractice insurance generally available but recommending development of contingency plans to ensure future availability).

36. FRANK A. SLOAN ET AL., INSURING MEDICAL MALPRACTICE 4 (1991); PAUL C. WEILER, MEDICAL MALPRACTICE ON TRIAL 1 (1991); *see, e.g.*, Tony Cunningham & Robin Lane, *Malpractice—the Illusory Crisis*, 54 FLA. B.J. 114, 114 (1980) (during 1974-75 Florida Medical Association raised over $800,000 for medical malpractice reform lobbying efforts).

37. Redish, *supra* note 34, at 760.

38. *Id.* at 760-61. No attempt will be made here to thoroughly review the events, causes, and impacts involved with the medical malpractice insurance crisis of the 1970s. The crisis has been reviewed extensively elsewhere. *See, e.g.*, HEW REPORT, *supra* note 35 (extensive empirical study of causes and impacts of crisis); TARKY LOMBARDI, JR., MEDICAL MALPRACTICE INSURANCE—A LEGISLATOR'S VIEW (1978) (reviewing malpractice crisis of 1970s); R. Scott Jenkins & Wm. C. Schweinfurth, Note, *California's Medical Injury Compensation Reform Act: An Equal Protection Challenge*, 52 SO. CAL. L. REV. 829, 848-52 (1979) (reviewing conflicting opinions on causes of crisis).

The insurance crisis, however, was not confined to the 1970s. Following the dramatic increases in costs in the mid-1970s, the growth of malpractice litigation and insurance costs has been cyclical, with periods of dramatic growth and periods of moderation and decline. *See* SLOAN et al., *supra* note 36, at 4-10; WEILER, *supra* note 36, at 2-7. Currently, the growth of litigation and costs is moderate, with some observers predicting a renewed problem of growth in the near future. *See, e.g.*, WEILER, *supra* note 36, at 2-7 (medical profession likely to face malpractice insurance crisis in 1990s).

It is necessary to note, however, that there is a long-running dispute over characterizing the events surrounding increasing malpractice litigation and insurance

Many state legislatures responded by passing a variety of proposals designed to limit the liability of hospitals, health care professionals, and insurers.[39] Among the medical malpractice reforms adopted in some states was a shortening of the limitation period for bringing medical malpractice actions.[40] The reasoning behind this particular reform was to reduce medical malpractice insurance costs by shortening the time period health care professionals would be subject to liability.[41]

---

costs as a "crisis." Historically, there has been substantial disagreement over the causes of the increase, its impact on the availability of health care, and whether a "crisis" even exists. *See, e.g.*, Jenkins & Schweinfurth, *supra*, at 831 n.2, 848-59 (reviewing "apparent" insurance crisis with focus on California events); *see also infra* notes 155-174 and accompanying text (discussing questionable extent of malpractice insurance crisis' impact on availability of health care and whether a "crisis" ever existed in Utah).

39. *See* Redish, *supra* note 34, at 761. By 1977, almost every state had commissioned studies or enacted legislation regarding the insurance crisis. *See* LOMBARDI, *supra* note 38, at 118-19.

40. *See* Redish, *supra* note 34, at 765.

41. *Id.* at 755. Prior to the medical malpractice crisis, malpractice actions were typically governed by statutes of limitation applicable to general tort or contract actions. *See* DAVID M. HARNEY, MEDICAL MALPRACTICE 247-48 (1973); *see also* LOUISELL & WILLIAMS, *supra* note 25, ¶ 13.01 (noting few states had statutes referring specifically to medical malpractice). Limitation periods also were often extended by judicial or statutory exceptions. Redish, *supra* note 34, at 765; WEILER, *supra* note 36, at 28. The two main exceptions to general limitation statutes were the general tolling provisions for legal disabilities and the "discovery" rule. *See* Redish, *supra* note 34, at 765. The discovery rule allows the applicable statute of limitation to be tolled in cases where the injury could not reasonably be discovered during the limitation period. *Id.*

These pre-crisis exceptions to statutes of limitation created relatively lengthy periods of potential liability that contributed to malpractice insurance costs in two ways. First, insurance companies were exposed to liability for substantial lengths of time. *Id.* Second, because malpractice rate-setting is "dependent on knowing with some degree of certainty the total potential losses for a policy year," insurance companies had to deal with substantial uncertainty in predicting losses for the long period of exposure to liability. *Id.*; HEW REPORT, *supra* note 35, at 41-42; PATRICIA M. DANZON, MEDICAL MALPRACTICE—THEORY, EVIDENCE & PUBLIC POLICY 174-76 (1985). Shorter statutes of limitation thus seek to reduce these costs by reducing uncertainty in ratemaking and the number of legitimate claims. Jenkins & Schweinfurth, *supra* note 38, at 956.

Additionally, the need to avoid stale malpractice claims by using statutes of limitation arguably is of particular importance due to evolving standards of care. Walter J. Wadlington, *Paying For Children's Medical Care: Interaction Between Family Law and Cost Containment*, 36 CASE W. RES. L. REV. 1190, 1216-17 (1986). For example, a health care provider may be disadvantaged by having to establish that the challenged medical procedure, since outmoded by new technology, met the standard of care that prevailed when the injury occurred. *Id.* This increases the provider's risk of being found negligent even though no fault existed at the time of injury. WEILER, *supra* note 36, at 28. It should be noted, however, that plaintiffs will to some extent suffer the same difficulties of stale and lost evidence that will burden a provider's defense. *See* Wadlington, *supra*, at 1217.

Such reasoning, of course, is particularly applicable to minors. Minors able to toll the limitation period until the age of majority potentially could subject health care professionals to liability for lengthy periods of time—a phenomenon sometimes referred to as a "long tail" problem.[42] As a result, some states shortened the limitation period for minors as well as for adults.[43] In almost all states where the limitation period was shortened for minors, the legislatures allowed at least a minimal tolling period for very young minors.[44] A typical modification requires a medical malpractice action to be brought within a set number of years after the minor has reached the age of six.[45] Apparently, the states using this approach reason that a minimum tolling period is necessary to protect very young minors' interests for a number of reasons. For example, such a minor may not be able to effectively communicate physical injuries to a parent or guardian.[46] In addition, less severe forms of infant brain damage may not be detectable or accurately evaluated until an infant matures.[47] Consequently, in almost all jurisdictions where state legislatures

---

42. See Redish, *supra* note 34, at 765; HEW REPORT, *supra* note 35, at 42.

43. *See* LOUISELL & WILLIAMS, *supra* note 25, ¶ 13.12.

44. *See infra* note 188 (surveying states shortening malpractice statute of limitation for minors).

45. *See, e.g.*, IND. CODE § 16-9.5-3-1 (1990) (action must be filed within two years "except that a minor under the full age of six (6) years shall have until his eighth birthday in which to file"); *see also* SPECIAL TASK FORCE ON PROFESSIONAL LIABILITY AND INSURANCE ACTION PLAN, AMERICAN MEDICAL ASSOCIATION, PROFESSIONAL LIABILITY IN THE '80S REPORT THREE 14 (1985) [hereinafter AMA TASK FORCE] (AMA advocates special statute of limitation for physicians who treat infants to bar claims unless brought by the age of six or eight).

46. *See, e.g.*, Johnson v. St. Vincent Hosp., 404 N.E.2d 585, 604 (Ind. 1980) (discussing rationale behind state legislature's tolling provision). According to the *Johnson* court, in determining the tolling period for minors in medical malpractice actions the Indiana Legislature

> may well have given consideration to the fact that most children by the time they reach the age of six years are in a position to verbally communicate their physical complaints to parents or other adults having a natural sympathy with them. Such communications and the persons whom they reach may to some appreciable degree stand surrogate for the lack of maturity and judgment of infants in this matter.

Id.; *see also* Douglas v. Stallings, 879 F.2d 1242, 1248 (7th Cir. 1989) (minimum tolling period presumably chosen to allow child to reach age that allowed child to communicate the existence and extent of any pain or injury).

47. *See* PAUL M. DEUTSCH & FREDERICK A. RAFFA, DAMAGES IN TORT ACTIONS § 133.30 (1992); Patricia H. Ellison, *Assessment of Brain Injury, in* MEDICAL MALPRACTICE: HANDLING OBSTETRIC AND NEONATAL CASES § 24, § 24.01 (Michael D. Volk & Melvin D. Morgan eds., 1986).

shortened the limitation period for minors, at least a minimal tolling period is allowed for very young minors.[48]

The Utah Legislature quickly responded to the health care and insurance industries' call for reform. In 1971, the Legislature shortened the statute of limitation applicable to medical malpractice claims from four years to two years and imposed a ten year statute of repose.[49] The Legislature acted again in 1976, passing the Utah Health Care Malpractice Act ("Malpractice Act").[50] The Malpractice Act mandated a four year statute of repose.[51] Al-

---

48. *See infra* note 188 (surveying states shortening malpractice statute of limitation for minors).

49. Act of Feb. 4, 1971, ch. 212, § 1, 1971 Utah Laws 614, 614 (amended 1976). Prior to 1971, adult plaintiffs were allowed four years to bring a cause of action. *See* Christiansen v. Rees, 436 P.2d 119, 200 (Utah 1968).

50. Ch. 23, 1976 Utah Laws 90 (codified as amended at UTAH CODE ANN. § 78-14-1 (1992)). In a statement of explicit findings, the Utah Legislature announced:

> The legislature finds and declares that the number of suits and claims for damages and the amount of judgments and settlements arising from health care has increased greatly in recent years. Because of these increases the insurance industry has substantially increased the cost of medical malpractice insurance. The effect of increased insurance premiums and increased claims is increased care cost, both through the health care providers passing the cost of premiums to the patient and through the provider's practicing defensive medicine because he views a patient as a potential adversary in a lawsuit. Further, certain health care providers are discouraged from continuing to provide services because of the high cost and possible unavailability of malpractice insurance.
>
> In view of these recent trends and with the intention of alleviating the adverse effects which these trends are producing in the public's health care system, it is necessary to protect the public interest by enacting measures designed to encourage private insurance companies to continue to provide health-related malpractice insurance while at the same time establishing a mechanism to ensure the availability of insurance in the event that it becomes unavailable from private companies.
>
> In enacting this act, it is the purpose of the legislature to provide a reasonable time in which actions may be commenced against health care providers while limiting that time to a specific period for which professional liability insurance premiums can be reasonably and accurately calculated; and to provide other procedural changes to expedite early evaluation and settlement of claims.

UTAH CODE ANN. § 78-14-2 (1992).

51. Utah Health Care Malpractice Act, ch. 23, § 4, 1976 Utah Laws 90, 93 (codified as amended at UTAH CODE ANN. § 78-14-4 (1992)). The statute of repose provides:

> (1) No malpractice action against a health care provider may be brought unless it is commenced within two years after the plaintiff or patient discovers, or through the use of reasonable diligence should have discovered the injury, whichever first occurs, but not to exceed four years after the date of the alleged act, omission, neglect or occurrence . . . .

though neither of these laws expressly applied to minors, apparently the Legislature intended to make the provisions applicable to minors as well as adults. When a Utah Supreme Court case subsequently called that view into question,[52] the Legislature amended the Malpractice Act in 1979 to expressly apply to minors.[53] In doing so, however, the Utah Legislature took the unique and unusually harsh position of eliminating any disability tolling provision for minors bringing medical malpractice actions.[54]

Like many other medical malpractice reforms, litigants in many jurisdictions have attacked legislative attempts to modify the application of general disability tolling statutes on constitutional grounds.[55] Plaintiffs have typically alleged that tolling modifications violate state and federal equal protection, due process, and open court constitutional provisions.[56] The results of these challenges have been mixed.[57] As discussed below, whether such treatment of minors is unconstitutional appears to depend substantially on the degree of scrutiny a court is willing to apply.

Id. The Act includes limited exceptions, however, for actions based on foreign objects wrongfully left in a patient's body, and cases involving fraudulent concealment by a health care provider. *Id.*

In addition to the statute of limitation change, Utah reforms included: a mandatory (but non-binding) claims screening panel, changes in the collateral source rule, prohibition of ad damnum clause in complaints, limits on contingency fees, periodic payment of damages, and provisions for voluntary, binding arbitration of claims. *See* UTAH CODE ANN. § 78-14-1 (1992).

52. *See infra* notes 71-83 and accompanying text (discussing the *Scott* decision).

53. Act of March 8, 1979, ch. 128, § 1, 1979 Utah Laws 739, 740 (codified as amended at UTAH CODE ANN. § 78-14-4(2) (1992)). This provision states: "The provisions of this section [78-14-4] shall apply to all persons, regardless of minority or other legal disability under [Utah's general disability tolling provision] or any other provision of law . . . ." UTAH CODE ANN. § 78-14-4(2) (1992).

54. *See infra* note 188 and accompanying text (surveying states shortening malpractice statute of limitation for minors).

55. *See* David R. Smith, *Battling a Receding Tort Frontier: Constitutional Attacks on Medical Malpractice Laws*, 38 OKLA. L. REV. 195, 200, 222 (1985).

56. *See Andrews, supra* note 23, at 472-86; *see generally* Jane M. Draper, Annotation, *Medical Malpractice Statutes of Limitation Minority Provisions*, 62 A.L.R. 4TH 758, 766-87 (1988) [hereinafter Annotation].

57. *See infra* notes 65-66 and accompanying text (citing cases with varying results in different jurisdictions).

IV. CONSTITUTIONAL SCRUTINY APPLIED TO STATUTES ELIMINATING TOLLING PROVISIONS FOR MEDICAL MALPRACTICE ACTIONS

A. Case Law From Other Jurisdictions

Under the United States Constitution, and under state constitutions providing the same level of protection, placing minors on equal footing with adults is not constitutionally prohibited.[58] Instead, it is within the power of a legislative body to do so as a matter of policy.[59] It is important to realize, however, that federal constitutional guarantees require courts to apply very little scrutiny to the abrogation of tolling provisions for minors bringing medical malpractice claims.

Medical malpractice reforms are commonly challenged under due process and equal protection provisions of the United States Constitution.[60] Although the issue has not been directly ad-

---

58. *See* Vance v. Vance, 108 U.S. 514, 521 (1883).

59. *Id.* The Supreme Court stated:

> [T]he Constitution of the United States . . . gives to minors no special rights beyond others, and it was within the legislative competency of the State of Louisiana to make exceptions in their favor or not. The exemptions from the operation of statutes of limitation, usually accorded to infants . . . do not rest upon any general doctrine of the law that they cannot be subjected to their action, but in every instance upon express language in those statutes giving them time after majority . . . to assert their rights.

Id.; *see also* Sherfey v. City of Brazil, 13 N.E.2d 568, 574 (Ind. 1938) (legislature not under constitutional mandate to suspend the obligation of statutes of limitation in the case of infancy or incapacity). For example, courts have consistently upheld the Federal Tort Claims Act's statute of limitation treating minors the same as adults when challenged under the federal constitution. *See, e.g.*, Robbins v. United States, 624 F.2d 971, 972 (10th Cir. 1980) ("[C]laimant's minority does not toll the running of the statute of limitation under Federal Tort Claims Act.").

Courts reviewing challenges to the Federal Tort Claims Act have rejected suggestions that the Act should protect minors from the failure or neglect of guardians to bring a minor's claim. *See* Zavala v. United States, 876 F.2d 780, 782 (9th Cir. 1989) (because parents have a duty to bring a child's claim, a child is bound by a parents decision not to bring an action and courts will not second guess their decision not to do so); *cf. Vance*, 108 U.S. at 521-22 (finding constitutional a law which did not toll limitation period for minors, but instead placed duty on proper officer of court and minor's guardian to act for minor). Courts applying the Act, however, at least have created an equitable tolling exception applied "where the injured party has no one with a legal duty to file suit on his behalf and the government, either by fraudulent concealment or negligent infliction of injury, has caused the plaintiff's inability to file a timely claim." *Zavala*, 876 F.2d at 783.

60. Smith, *supra* note 55, at 200-01.

dressed by the United States Supreme Court, the majority of courts applying federal constitutional principles—or applying state constitutional principles using a federal analysis—scrutinize these particular issues using undemanding standards.[61] For example, under an equal protection analysis, the standard commonly applied under federal precedent is a "rational basis" standard, which merely requires that the challenged statute rationally promote a legitimate governmental objective.[62] Similarly, under a due process analysis, federal precedent requires a legislative act to be sustained if not "arbitrary and irrational."[63] Both standards have generally proven deferential to the choices made by legislatures and rarely result in a statute being overturned.[64] Accordingly, when applying minimal scrutiny standards to the modification of tolling provisions for minors in medical malpractice actions, a majority of jurisdictions have upheld such legislation as constitutional.[65]

---

61. *Id.* at 208-09 (most courts review medical malpractice reforms under rational relation test). *See generally* Annotation, *supra* note 56, at 766-82 (reviewing cases involving constitutional challenges to statute of limitation reforms for minors bringing malpractice actions).

62. Usery v. Turner Ellchorn Mining Co., 428 U.S. 1, 15 (1976); *see also* Smith, *supra* note 55, at 202 (legislative act upheld if "not wholly arbitrary and capricious") (citing Minnesota v. Clover Leaf Creamery Co., 449 U.S. 456 (1981)). The rational basis standard is the least demanding of three standards; two standards, one "intermediate" and the other "strict" scrutiny subject challenged legislation to a more rigorous standard of review. *Id.* at 202-04. Strict scrutiny is reserved for classifications affecting a fundamental right, such as voting, or suspect classifications such as race. Massachusetts Bd. of Retirement v. Murgia, 427 U.S. 307, 312 (1976) ("[E]qual protection analysis requires strict scrutiny of a legislative classification only when the classification impermissibly interferes with the exercise of a fundamental right or operates to the peculiar disadvantage of a suspect class.") (footnotes omitted). The modification of minors' rights by eliminating a statute of limitation tolling provision is unlikely to result in strict scrutiny. Smith, *supra* note 55, at 204; *see, e.g.*, Hargett v. Limberg, 598 F. Supp. 152, 157 (D. Utah 1984) (rejecting strict scrutiny as the appropriate standard for reviewing constitutionality of malpractice reform). *But see* White v. State, 661 P.2d 1272, 1274-75 (Mont. 1983) (treating recovery of damages for bodily injuries as fundamental right requiring strict scrutiny).

63. Smith, *supra* note 55, at 210. As with equal protection analysis, a due process legislative challenge may require more rigorous scrutiny under circumstances involving fundamental rights or restrictions on political processes. *Id.* Most courts reject applying this greater scrutiny to medical malpractice reforms, however. *Id.* at 212.

64. *See* Redish, *supra* note 34, at 769-82.

65. *See* Smith, *supra* note 55, at 212; WEILER, *supra* note 36, at 39 (Federal Constitution offers little or no relief for plaintiffs dissatisfied with malpractice reform); *see, e.g.*, Douglas v. Stallings, 870 F.2d 1242, 1249-50 (7th Cir. 1989) (rejecting constitutional challenge under Indiana law); Reese v. Rankin Fite Memorial Hosp., 403 So. 2d 158, 161 (Ala. 1981) (rejecting equal protection challenge); Rohrabaugh v.

In contrast, jurisdictions using a heightened form of scrutiny, or applying a different analysis under state constitutional provisions, have often found modification of tolling provisions for minors unconstitutional.[66] Thus, a crucial issue in evaluating the constitutionality of Utah's treatment of minors regarding medical malpractice actions is determining what kind of scrutiny Utah courts would apply in evaluating that treatment.

B. Utah's Equal Protection and Open Court Guarantees

Utah is among those states that have developed a body of state constitutional law distinct from federal constitutional law.[67] More specifically, using an independent state constitutional analysis, a majority of the Utah Supreme Court recently demonstrated a willingness to substantially scrutinize legislation infringing on the right to seek a remedy for wrongful or negligent

---

Wagoner, 413 N.E.2d 891, 894-95 (Ind. 1980) (same); Johnson v. St. Vincent Hosp., Inc., 404 N.E.2d 585, 604-05 (Ind. 1980) (same); Wheeler v. Lenski, 658 P.2d 1056, 1059 (Kan. Ct. App. 1983) (same); Maine Medical Ctr. v. Cote, 577 A.2d 1173, 1176-77 (Me. 1990) (same); Cioffi v. Guenther, 370 N.E.2d 1003, 1006 (Mass. 1977) (same); Jones v. Salem Hosp. 762 P.2d 303, 308-09 (Or. Ct. App. 1988) (same); Halverson v. Tydrich, 456 N.W.2d 852, 858 (Wis. Ct. App. 1990) (same). *But see* Lyons v. Lederle Lab., 440 N.W.2d 769, 771 (S.D. 1989) (violates equal protection guarantee under rational basis standard).

66. Smith, *supra* note 55, at 204-07; *see, e.g.*, Barrio v. San Manuel Div. Hosp., 692 P.2d 280, 286 (Ariz. 1984) (medical malpractice limitation period applicable to minors unconstitutional under state guarantee of right to recover for injury); Strahler v. St. Luke's Hosp., 706 S.W.2d 7, 11-12 (Mo. 1986) (medical malpractice limitation period unconstitutionally applied to minors, violating guaranteed access to courts); Carson v. Maurer, 424 A.2d 825, 830, 833-34 (N.H. 1980) (medical malpractice limitation period violates equal protection guarantee under intermediate scrutiny standard); Mominee v. Scherbarth, 503 N.E.2d 717, 723 (Ohio 1986) (medical malpractice limitation period violates due process guarantee of state constitution); Sax v. Votteler, 648 S.W.2d 661, 667 (Tex. 1983) (medical malpractice limitation period violates state open courts guarantee); *Lyons*, 440 N.W.2d at 771 (medical malpractice limitation period violates equal protection guarantee under rational basis standard).

67. *See* Mountain Fuel Supply v. Salt Lake City, 752 P.2d 884, 889-90 (Utah 1988); Milo S. Marsden, Comment, *The Utah Supreme Court and the Utah State Constitution*, 1986 UTAH L. REV. 319, 319-21.

A state's construction and application of state constitutional provisions is not controlled by the federal courts' construction and application of the federal constitution. *See, e.g.*, Malan v. Lewis, 693 P.2d 661, 670 (Utah 1984) (federal precedent regarding federal equal protection clause not binding on application of state equal protection guarantee as long as result does not violate the Federal Constitution). This trend among states to pursue independent state constitutional jurisprudence is well-recognized and has generated a considerable body of literature. *See Mountain Fuel, supra* at 889-90; *Andrews, supra* note 23, at 472-73.

personal injury.[68] The unique, independent analysis applied in recent cases interpreting the Utah Constitution will likely guide the court as it considers the Malpractice Act's treatment of minors. This is particularly true since earlier case law addressing the disability tolling provision and the Malpractice Act's statute of limitation send conflicting signals on the constitutionality of the Malpractice Act's treatment of minors.

1. Utah's Historical Treatment of Statutes of Limitation Regarding Minors and Medical Malpractice Actions

Utah has long recognized that although statutes of limitation are a matter of policy properly decided by the legislature,[69] they must operate within constitutionally permissible limits.[70] In addition, the Utah Supreme Court has previously discussed the constitutionality of eliminating the disability tolling provision, and ruled on the constitutionality of the Malpractice Act's two year statute of limitation as it applies to adults. The court's opinions, however, do not clearly indicate how the court would currently evaluate the Malpractice Act's treatment of minors.

For example, in *Scott v. School Board*,[71] the Utah Supreme Court reviewed a provision of the Utah Governmental Immunity Act which required plaintiffs to give notice of a claim against the state within a certain time period.[72] The minor plaintiff argued that the notice requirement was tolled by operation of the general tolling statute even though the notice provision of the Immunity Act did not reference any exceptions.[73] The court expressly rejected prior decisions holding that the tolling provision was not

68. *See, e.g.*, Condemarin v. University Hosp., 775 P.2d 348, 356 (Utah 1989) (majority of court applying some form of heightened scrutiny to legislation infringing on right to bring cause of action for negligent personal injury).

69. *See, e.g.*, Gray Realty Co. v. Robinson, 184 P.2d 237, 247 (Utah 1947) (statute of repose enacted by legislature as a matter of policy).

70. *See, e.g.*, Toronto v. Sheffield, 222 P.2d 594, 596 (Utah 1941) ("The Legislature may bar a claim within a reasonable time within the effective date of a statute enacted for that purpose, but may not constitutionally bar such claim without allowing some time to elapse during which claimant may bring an action thereon after the effective date of the statute.").

71. 568 P.2d 746 (Utah 1977).

72. *Id.* at 747.

73. *Id.* at 746-47.

applicable in such cases,[74] and instead held that the disability tolling provision automatically applied.[75] The court reasoned that the general tolling statute, and a recent amendment applying the tolling statue to a similar notice provision, suggested a legislative intent to protect the causes of minors.[76]

Significantly, the *Scott* court also noted that minors are not able to bring an action on their own behalf and must rely on parent or guardians instead.[77] Moreover, the court emphasized that

> parents, or natural guardians, have no specific legal duty to perform and have no responsibility to their minor off-spring other than their moral obligation. Consequently, in matters of this kind, when a parent . . . fails for one reason or another to give notice, file suit, or otherwise protect the minor's legal interests, the minor is left completely without a remedy. This was undoubtedly one of the prime considerations which prompted the legislature to toll the statute during the minority of a claimant [in the similar notice provision].[78]

The court concluded that minor claimants are entitled to tolling provisions "in all cases" because "[t]o hold otherwise is a denial of due process and equal protection."[79]

Arguably, the *Scott* case stands for the broad proposition that failure to apply the general tolling provision is unconstitutional "in all cases."[80] The precise effect of *Scott* is not so clear, however.

---

74. *See* Varoz v. Sevey, 506 P.2d 453 (Utah 1973); Gallegos v. Midvale City, 492 P.2d 1335 (Utah 1972).

75. *Scott*, 568 P.2d at 748.

76. According to the *Scott* court, the legislature's "reason for not [providing tolling for minors] in governmental immunity cases as we are faced with here is entirely unclear. However, the general legislative intent to protect the causes of minors is abundantly clear by [the recent amendment] and the specific provisions of the general statute of limitations." *Id.* at 747-48.

77. *Id.* at 747 (emphasizing minor's inability to give or receive lawful notice because minority status confers no legal standing).

78. *Id.*

79. *Id.* at 748. The *Scott* case ultimately provided the basis for applying the tolling provision to two other statutes of limitation that could have been interpreted as treating minors as adults. *See* Szarak v. Sandoval, 636 P.2d 1082, 1085 (Utah 1981) (applying tolling provision to minor's paternity and child support action); Switzer v. Reynolds, 606 P.2d 244, 249 (Utah 1980) (applying tolling provision to a minor's action for wrongful death).

80. *Scott*, 568 P.2d at 748. This position is supported by Justice Howe's characteriza-

The Utah Supreme Court recently noted that there was a question whether the holding of *Scott* was constitutionally based, concluding that *Scott* "at least" stands for the principle that absent specific intent, limitation periods are tolled during minority.[81] This suggests that despite the *Scott* court's explicit language, the current Utah Supreme Court might view *Scott* as resting on statutory construction and legislative intent principles instead of constitutional foundations. This view is based on the rule of constitutional jurisprudence that a court should avoid resolving an issue as a constitutional question if the court can resolve it by any other means.[82] Therefore, it is unclear whether the *Scott* case would be dispositive of the question posed by the Malpractice Act's treatment of minors.[83] Nevertheless, the case is at least strongly persuasive that the Malpractice Act's elimination of the tolling provision is unconstitutional.

Contrasting the *Scott* decision is the 1981 case of *Allen v. Intermountain Health Care, Inc.*[84] In *Allen*, the supreme court evaluated an equal protection challenge to the Malpractice Act's

---

tion of the *Scott* case in a subsequent case addressing the judicial application of the discovery rule in wrongful death actions. Myers v. McDonald, 635 P.2d 84, 88 (Utah 1981) (Howe, J., concurring). According to Justice Howe, the *Scott* court recognized that the denial of tolling provisions for minors may be a denial of constitutional rights. *Id.*

81. Blum v. Stone, 752 P.2d 898, 900 & n.2 (Utah 1988) (citing Szarak v. Sandoval, 636 P.2d 1082 (Utah 1981)); Switzer v. Reynolds, 606 P.2d 244, 250 n.8 (Utah 1980) (Hall, J., dissenting)).

The Utah Legislature reacted to the *Scott* decision as if it was merely a case of statutory interpretation. Prior to *Scott*, the Malpractice Act generally provided for only a two year statute of limitation for all medical malpractice claims. *See* Utah Health Care Malpractice Act, ch. 23, 1976 Utah Laws 90 (codified as amended at UTAH CODE ANN. § 78-14-4(2) (1992)). The legislature amended the provision to "overturn" *Scott* and to clarify its intention to eliminate the general tolling provision in medical malpractice cases: "'[T]his section shall apply to all persons, regardless of minority or other legal disability under § 78-12-36 or any other provision of the law . . . .'" *Blum*, 752 P.2d at 900 n.2. (quoting UTAH CODE ANN. § 78-14-4(2) (1987)). Of course, if the *Scott* decision is constitutionally based, the Legislature would be unable to "overturn" the decision by legislative act. *See id.* at 901.

82. *See, e.g.*, Millett v. Clark Clinic Corp., 609 P.2d 934, 936 (Utah 1980) (court's responsibility and purpose in interpreting statutes is to give effect to legislative intent); Hoyle v. Monson, 606 P.2d 240, 242 (Utah 1980) ("fundamental rule of long-standing is that unnecessary [constitutional] decisions are to be avoided and that the courts should pass upon the constitutionality of a statute only when such a determination is essential to the decision in a case"); *see also* Brief for Appellee at 13, Blum v. Stone, 752 P.2d 898 (Utah 1988) (No. 20288) (asserting *Scott* is not of constitutional dimension).

83. *See Blum*, 752 P.2d at 901 (leaving open question whether *Scott* was of constitutional dimension).

84. 635 P.2d 30 (Utah 1981).

statute of limitation as it applied to an adult.[85] The *Allen* court found the Malpractice Act's limitation period constitutional under the undemanding "rational basis" test.[86] In a brief opinion reviewing the Malpractice Act's stated purposes,[87] Chief Justice Hall summarily determined:

> The legislature exercised its discretionary prerogative in determining that the shortening of the statute of limitation . . . would insure the continued availability of adequate health care services. In the absence of a showing to the contrary, we conclude that the legislature's determination is not so arbitrary or unreasonable as to exceed constitutional prohibitions.[88]

The outcome and the analysis applied in *Allen* strongly suggest that the supreme court deems the Malpractice Act's elimination of tolling provisions constitutional. *Allen* also illustrates the deference courts afford to legislation analyzed under the "rational basis" test. At first glance, *Allen* might lead to the conclusion that the supreme court would similarly find the Malpractice Act's treatment of minors acceptable. A number of significant factors argue against such a conclusion, however. First, *Allen* addresses the Malpractice Act's statute of limitation as it applies generally to adults and not to minors specifically.[89] As previously noted, minors are in a distinctly different position because they lack standing to bring an action on their own behalf.[90] In addition, the *Allen* court addressed the constitutionality of the statute solely under an equal protection analysis.[91] The statute was not challenged under the Open Courts Provision, a provision which has subsequently proven fatal to a number of other statutes of repose.[92] Moreover, as argued below, since the *Allen* decision the supreme court has shown more willingness to

---

85. *Id.* at 31.
86. *Id.*
87. *See supra* note 50 (quoting text of legislative findings).
88. *Allen*, 635 P.2d at 32.
89. *Id.* at 30-31.
90. *See supra* notes 2-5 and accompanying text (discussing minors' legal status).
91. *Allen*, 635 P.2d at 30-32.
92. *See infra* notes 95-105 and accompanying text (discussing case declaring statute of repose unconstitutional under Utah's Open Courts Provision).

carefully scrutinize the constitutionality of legislation affecting the right to recover for personal injury.[93] Consequently, the *Allen* decision is not likely to be as influential as it appears at first glance.[94]

The *Scott* and *Allen* decisions, as Utah's precedent most closely related to the question of the constitutionality of the Malpractice Act's treatment of minors, send mixed messages. Review of more recent case law elaborating on Utah's unique constitutional requirements, however, suggests that the Malpractice Act will be subject to significantly greater scrutiny than was employed in *Allen*.

2. Utah's Current Open Courts Analysis

In the recent case of *Berry v. Beech Aircraft Corp.*,[95] the Utah Supreme Court clearly established a substantial and independent state constitutional protection providing certain rights under the Open Courts Provision of the Utah Constitution.[96] According to the *Berry* court, this provision, for which there is no federal constitutional analogue, "guarantees access to the courts and a judicial procedure that is based on fairness and equality."[97] The Provision does not create or absolutely preserve any vested rights under the law. Rather, it protects the arbitrary infringe-

---

93. *See infra* notes 106-36 and accompanying text (discussing standard applied in recent equal protection cases).

94. The Federal District Court for Utah has specifically addressed the constitutionality of the Malpractice Act's treatment of minors. In Hargett v. Limberg, 598 F. Supp. 152 (D. Utah 1984), the court applied the same equal protection test as *Allen*, concluding that eliminating tolling provisions in medical malpractice claims was constitutional. *Id.* at 157-58. While persuasive, this case is not binding on Utah courts and, more significantly, it was decided prior to more recent cases which clearly apply a more heightened level of scrutiny. *See infra* notes 106-36 and accompanying text (discussing standard applied in recent equal protection cases).

95. 717 P.2d 670 (Utah 1985).

96. *See id.* at 674-83. The Open Courts Provision provides:

> All courts shall be open, and every person, for an injury done to him in his person, property, or reputation, shall have remedy by due course of law, which shall be administered without denial or unnecessary delay; and no person shall be barred from prosecuting or defending before any tribunal in this state, by himself or counsel, any civil cause to which he is a party.

UTAH CONST. art. 1, § 11; *see also Berry*, 717 P.2d at 674 (referring to provision as "open courts provision").

97. *Berry*, 717 P.2d at 675.

ment or abrogation of existing rights.[98] Thus, the Open Courts Provision does not prevent the legislature from abolishing a particular right.[99] Instead, the protection provided by the clause is balanced against the necessary power of the legislature to change and update the law as conditions require.[100] Legislative acts which unacceptably impair litigants' rights, however, are unconstitutional.[101] Accordingly, the court must substantially scrutinize any legislation infringing on existing rights protected by the Open Courts Provision.[102]

In line with this view of the Open Courts Provision, the *Berry* court devised a two-part test for evaluating a legislative abrogation or restriction of a right protected under the Open Courts Provision:

---

98. *Id.* at 675-76. "[O]nce a cause of action under a particular rule of law accrues to a person by virtue of an injury to his rights, that person's interest in the cause of action and the law which is the basis for a legal action becomes vested, and a legislative repeal of the law cannot constitutionally divest the injured person of the right to litigate the cause of action to a judgment." *Id.* at 676. Furthermore, a legislature could not constitutionally abolish all right of action for injury to one's person. *Id.* at 676 n.3.

99. *Id.* at 676-77.

100. *Id.* The Open Courts Provision has a strong but somewhat unclear relationship to the Utah Constitution's Due Process Clause: the provisions are "related both in their historical origins and to some extent in their constitutional functions. To a degree, the two provisions are complementary and even overlap, but they are not wholly duplicative. Both act to restrict the powers of both the courts and legislature." *Id.* at 675. Nevertheless, the balancing test derived from the Open Courts Provision in *Berry* has been referred to as a "due process" test. *See Condemarin*, 775 P.2d at 360, 367 (Zimmerman, J., concurring). *But see id.* at 369-70 (Stewart, J., concurring separately) (objecting to Justice Durham's and Justice Zimmerman's blurring due process, equal protection, and open courts analyses). For purposes of this Comment, the balancing test derived from the Open Courts Provision in *Berry* will be referred to as the "*Berry* test."

101. *Berry*, 717 P.2d at 676-77.

102. *See id.* In *Condemarin*, Justice Zimmerman expanded on the rationale underlying the Open Courts Provision:

> The constitution's drafters understood that the normal political processes would not always protect the common law rights of all citizens to obtain remedies for injuries. At any one time, only a small percentage of the citizenry will have recently been harmed and therefore will need to obtain a remedy from the members of any particular defendant class. The vast majority of the populace will have no interest in opposing legislative efforts to protect such a defendant class because the majority will not readily identify with those few persons unlucky enough to have been harmed. And those few persons directly affected will, in all likelihood, lack the political power to prevent the passage of legislation that, in essence, requires every member of the citizenry who is injured by members of the defendant class to bear some or all of the cost of those injuries.

Condemarin, 775 P.2d at 367 (Zimmerman, J., concurring) (citations omitted).

> First, . . . the law [must] provide an injured person an effective and reasonable alternative remedy "by due course of law" for vindication of his constitutional interest. The benefit provided by the substitute must be substantially equal in value or other benefit to the remedy abrogated in providing essentially comparable substantive protection to one's person, property, or reputation, although the form of the substitute remedy may be different.
>
> . . . .
>
> Second, if there is no substitute or alternative remedy provided, abrogation of the remedy or cause of action may be justified only if there is a clear social or economic evil to be eliminated and the elimination of an existing legal remedy is not an arbitrary or unreasonable means for achieving the objective.[103]

---

103. *Berry*, 717 P.2d at 680. Of course, as a threshold matter, it is necessary to determine whether the challenged legislation does indeed affect a right protected under the Open Courts Provision. *See Condemarin*, 775 P.2d at 368 (Zimmerman, J., concurring) (burden of proof shifts "once it is shown that the enactment under scrutiny does, in fact, infringe upon the interests enumerated in article I, section 11").

As a general rule, challenged legislation is accorded a presumption of constitutionality. *See infra* note 110 (discussing basic principles of constitutional analysis). At least two Justices, however, have suggested that in analyzing legislation infringing on rights protected under the Open Courts Provision, the presumption of constitutionality would be reversed and the burden of meeting the test would fall on the proponents of the challenged law. *See* Horton v. Goldminer's Daughter, 785 P.2d 1087, 1096 (Utah 1989) (Zimmerman, J. & Durham, J., concurring). Nonetheless, it does not appear that a majority of the court would adhere to this view of shifting the burden of proof. *See id.* at 1096 (Justice Stewart, Associate Chief Justice Howe, and Chief Justice Hall apparently not accepting Justice Durham's and Justice Zimmerman's concurring opinion shifting burden of proof).

On the other hand, it does appear that a majority of the court would at least reject the presumption of constitutionality normally accorded legislation and independently analyze the basis for the legislation. In his equal protection opinion in *Condemarin*, Justice Stewart stated that "to presume the constitutionality of a statute when the statute deprives one of a right established by Article I, section 11 of the state constitution is to fail to give any greater weight to a constitutional right than to a nonconstitutional interest, such as a general social or economic interest." *Condemarin*, 775 P.2d at 370 (Stewart, J., concurring). Thus, it appears that Justice Stewart would eliminate the presumption of constitutionality when faced with the infringement of a right protected by the Open Courts Provision. Justice Stewart's view, combined with those of Justices Durham and Zimmerman, appears to form a majority of the court that would at least eliminate the presumption of constitutionality. This group would likely make its own in-depth inquiry into the facts supporting justification when faced with a law that infringes on a right protected by the Open Courts Provision. *Cf.* LeCroy v. Hanlon, 713 S.W.2d 335, 341 (Tex. 1986) (proponent of legislation carries burden of proof in balancing portion of open courts test).

Under this test, the *Berry* court held unconstitutional a statute of repose limiting the time allowed for bringing a products liability claim.[104] Subsequently, the *Berry* test was affirmed and served as the basis for striking down two other statutes of repose.[105] Thus, the Utah Open Courts Provision has been interpreted as a mandate to the courts to substantially scrutinize legislation infringing on the right to seek a judicial remedy for personal injury. As such, the *Berry* test serves as one model for analyzing the constitutionality of the Medical Malpractice Act's elimination of the disability tolling provision.

3. *Utah's Current Equal Protection Analysis*

As with its open courts analysis, the Utah Supreme Court appears willing to substantially scrutinize legislation under state equal protection principles in circumstances requiring only minimal scrutiny under a federal constitutional analysis. Article I, section 24 of the Utah Constitution provides: "All laws of a general nature shall have uniform operation."[106] Although the language is different, this clause is regarded as Utah's equivalent of the Federal Constitution's fourteenth amendment equal protection guarantee.[107] Historically, when applying the state Equal Protection Provision, the Utah Supreme Court has relied on an analytical structure provided by federal precedent.[108] Indeed, the court has stated that the state and federal provisions are

---

104. *Berry*, 717 P.2d at 683. The statute required a products liability claim to be brought within six years after the date of initial purchase of the product, or ten years after its manufacture. *Id.* at 671-72. Citing a number of factors, the *Berry* court concluded that the statute was unreasonable, arbitrary, and would not further the statutory objectives. *Id.* at 681-83.

105. Horton v. Goldminer's Daughter, 785 P.2d 1087, 1094 (Utah 1989) (invalidating architects and builders statute of repose); Sun Valley Water Beds v. Herm Hughes & Son, 782 P.2d 188, 192-94 (Utah 1989) (same); Wrolstad v. Industrial Comm'n, 786 P.2d 243, 245 (Utah Ct. App. 1989) (invalidating statute of repose contained in Utah Occupational Disease Disability Law).

106. UTAH CONST. art. I, § 24.

107. Amax Magnesium Corp. v. Utah State Tax Comm'n, 796 P.2d 1256, 1261 (Utah 1990); Malan v. Lewis, 693 P.2d 661, 669 n.13 (Utah 1984). *Compare* UTAH CONST. art. I, § 24 ("All laws of a general nature shall have uniform operation.") *with* U.S. CONST. amend. XIV, § 1, cl.2 ("No state shall . . . deny to any person within its jurisdiction the equal protection of the laws.").

108. *See* Mountain Fuel Supply Co. v. Salt Lake City Corp., 752 P.2d 884, 888 (Utah 1988).

considered equivalent in many circumstances.[109] Like other state courts relying on federal principles, the Utah Supreme Court has afforded broad deference to legislative classifications when applying the state Equal Protection Provision to economic regulations.[110]

A number of recent Utah cases, however, have made clear that the court may review challenged legislation under an equal protection standard "which is at least as exacting and, in some circumstances, more rigorous than the standard applied under the federal constitution."[111] Beginning with the 1984 case of *Malan v. Lewis*,[112] the Utah Supreme Court indicated that its analysis of the state equal protection guarantee is not controlled by precedent developed under the Federal Constitution.[113] While such precedent may be persuasive, "[t]he different language of Article I, § 24, the different constitutional contexts of the two provisions, and different jurisprudential considerations may lead

---

109. *See id.*; Baker v. Matheson, 607 P.2d 223, 243-45, 243 n.4 (Utah 1979).

110. *See* Blue Cross & Blue Shield v. State, 779 P.2d 634, 637 (Utah 1989); City of W. Jordan v. Utah State Retirement Bd., 767 P.2d 530, 537 (Utah 1988); *Mountain Fuel*, 752 P.2d at 888.

As recently outlined in *Blue Cross*, the concept underlying the Utah Equal Protection Provision entails

> "the settled concern of the law that the legislature be restrained from the fundamentally unfair practice" of classifying persons in such a manner that those who are similarly situated with respect to the purpose of a law are treated differently by that law, to the detriment of some of those so classified. In scrutinizing a legislative measure under article I, § 24, we must determine whether the classification is reasonable, whether the objectives of the legislative action are legitimate, and whether there is a reasonable relationship between the classification and the legislative purposes.

Blue Cross, 779 P.2d at 637 (citation omitted).

This analysis operates under the well-established principle that "all statutes are presumed to be constitutional and the party challenging a statute bears the burden of proving its invalidity." *Id.* Furthermore, using broad deference in reviewing economic legislation leads the court to sustain the legislation if "'facts can reasonably be conceived which would justify the distinctions or differences in state policy [expressed by the challenged legislation] as between different persons.'" *Id.* (quoting *Baker*, 607 P.2d at 244). The court does not, however, accept any conceivable legislative reason for the enactment; rather, it "judge[s] such enactments on the basis of reasonable or actual legislative purposes.'" *Id.* (quoting *Malan*, 693 P.2d at 671 n.14).

111. *Mountain Fuel*, 752 P.2d at 889; *see Blue Cross*, 779 P.2d at 637; *Malan*, 693 P.2d at 671; *see also* Greenwood v. City of N. Salt Lake, 817 P.2d 816, 821 (Utah 1991) (state test is somewhat more restrictive than federal test).

112. 693 P.2d 661 (Utah 1984).

113. *Id.* at 670.

to a different result in applying [Utah] equal protection principles than might be reached under federal law."[114]

Following this view, the court has diverged from federal precedent and, in certain instances, applied more than minimal scrutiny to legislation challenged on an equal protection basis. In *Malan*, for example, the court invalidated an automobile guest statute[115] using an equal protection analysis that was later characterized as allowing greater protection to individuals than that allowed under federal law.[116] The *Malan* opinion, however, left unclear when and to what degree a heightened level of scrutiny would be applied. The *Malan* court expressly denied that the scrutiny it applied rose to the level characterized as "intermediate" under federal principles.[117] However, the court implied that some level of increased equal protection scrutiny was justified in *Malan*, partially because the challenged legislation infringed on an individual's "bodily integrity," a right separately protected by the Utah Open Courts Provision.[118] Nevertheless, the case left many unanswered questions regarding the contours of its analysis.[119]

In *Condemarin v. University Hospital*,[120] the court partially clarified the ambiguities left by *Malan*. It indicated that a majority would apply some heightened level of equal protection scrutiny to legislation impairing a right protected under the Open Courts Provision. In *Condemarin*, a plurality of the Utah Supreme Court held unconstitutional a section of the Utah Governmental Immunity Act[121] that limited the amount of a plaintiff's recovery,

---

114. *Id.*

115. *Id.* at 675.

116. *See* Condemarin v. University Hosp., 775 P.2d 348, 369, 373 (Utah 1989) (Stewart, J., concurring).

117. *Malan*, 693 P.2d at 674 n.17.

118. *Id.*

119. As an indication of the ambiguity left by *Malan*, compare *Condemarin*, 775 P.2d at 369 (Stewart, J., concurring) (*Malan* allows greater protection under Utah constitution than provided by Federal Constitution) with *id.* at 379-80 (Hall, C.J., dissenting) (standards used by Utah Supreme Court and federal courts do not conflict and "no attempt will be made to differentiate between state & federal equal protection arguments").

120. 775 P.2d 348 (Utah 1989). The court's opinion consisted of a plurality opinion by Justice Durham, a separate concurring opinion by Justice Zimmerman, a separate concurring opinion by Justice Stewart, and a dissenting opinion by Chief Justice Hall, in which Associate Chief Justice Howe concurred.

121. UTAH CODE ANN. § 63-30-1 (1989 and Supp. 1991).

as applied to the University of Utah Hospital.[122] The four *Condemarin* opinions focused on selecting the proper analysis to determine constitutionality, with three of the opinions ultimately relying on different tests and standards. Nevertheless, the plurality and concurring opinions in *Condemarin* can be viewed as supporting the proposition that a plaintiff's right to recover for a tortious injury is constitutionally protected to an extent that mandates a heightened level of scrutiny under Utah's Equal Protection Provision.

Justice Durham analyzed the issue under both equal protection and due process approaches.[123] She concluded that the latter approach, embodying a balancing test, was the more appropriate analytical vehicle to decide the case before her because it was more flexible and straightforward.[124] Arguing that the *Berry* test embodies such a due-process-type balancing analysis, she applied *Berry* and concluded that part of the statute at issue was unconstitutional.[125] Nevertheless, in discussing the appropriate equal protection analysis of legislation dealing with the individual's right to seek compensation for harm from a wrongful or negligent act, Justice Durham expressly rejected the notion of granting "almost total deference" to the legislature accorded under the federal "mere rationality" test.[126] Instead, Justice Durham concluded that she would apply a "heightened standard of review."[127]

Justice Zimmerman concurred with Justice Durham's *Berry* analysis.[128] Like Justice Durham, Justice Zimmerman concluded that the *Berry* test provided a more rational means of analyzing

---

122. *Condemarin*, 775 P.2d at 366. For a brief overview of the entire *Condemarin* opinion, see Scot D. Tucker, Recent Development, *Constitutionality of Damage Recovery Limitation Caps for Governmentally Operated Health Care Facilities*, 1990 UTAH L. REV. 143.

123. *Condemarin*, 775 P.2d at 352-64 (plurality opinion).

124. *Id.* at 357 (plurality opinion).

125. *Id.* at 352-64 (plurality opinion).

126. *Id.* at 354 (plurality opinion). Although she characterized the right to a remedy for personal injury as a "fundamental principle," Justice Durham stated that this principle did not rise to the level of a fundamental right for purposes of equal protection analysis. *Id.* at 354, 359 (plurality opinion).

127. *Id.* at 356 (plurality opinion). Justice Durham did not specify the level of scrutiny she would apply. She did, however, favorably refer to cases from other jurisdictions that applied "a real and thoughtful examination of legislative purpose and the relationship between the legislation and that purpose." *Id.* (plurality opinion).

128. *Id.* at 366 (Zimmerman, J., concurring).

the rights protected under the Open Courts Provision than the "rigid" equal protection analysis.[129] He likewise discussed the Equal Protection Provision, and found "little reason why the analytical framework used to test the constitutionality of legislation under [Utah's Equal Protection Provision] must ape the rigid two- (or three-) level analysis of the federal equal protection cases."[130] Furthermore, in addressing the level of scrutiny required for legislation infringing on rights protected under the Open Courts Provision, Justice Zimmerman concluded that a standard higher than that accorded ordinary economic legislation is appropriate.[131] Consequently, if required to apply an equal protection analysis, Justice Zimmerman, like Justice Durham, would likely opt for a heightened level of scrutiny when addressing legislation infringing the right to bring a personal injury action.

Justice Stewart, on the other hand, applied an equal protection analysis using only a level of scrutiny with "bite."[132] Relying on *Malan v. Lewis*,[133] Justice Stewart characterized the appropriate standard to be used when evaluating an "important" right protected under the Open Courts Provision: "The statutory classifications must be reasonable, and the statute that creates the classification must in fact reasonably and substantially further the legislative purpose."[134] Because the legislation at issue involved

---

129. *Id.* at 367 (Zimmerman, J., concurring). According to Justice Zimmerman, applying the *Berry* test avoids "being bound into the analytical straitjacket that has been fashioned out of the federal equal protection clause for 'fundamental' rights and the tempting parallel construction of the Utah Constitution's uniform-operation-of-the-laws provision." *Id.* (Zimmerman, J., concurring).

130. *Id.* (Zimmerman, J., concurring).

131. *Id.* at 368 (Zimmerman, J., concurring).

132. *Condemarin*, 775 P.2d at 373 (Stewart, J. concurring). Justice Stewart found no need to rely on a due process analysis since it had not been raised by the parties. *Id.* at 369 (Stewart, J. concurring). He also disagreed with Justice Durham's and Justice Zimmerman's meshing the *Berry* open courts analysis with an analysis called for under the Utah Constitution's Due Process Clause. *Id.* (Stewart, J. concurring).

133. 693 P.2d 661 (Utah 1984).

134. *Condemarin*, 775 P.2d at 373 (Stewart, J., concurring) (citation omitted). Justice Stewart further noted:

> The appropriate standard . . . has more bite than the minimum scrutiny standard but does not purport to require the Legislature to find the least restrictive manner of furthering its purpose. But neither does it allow, on the other hand, such wide latitude as to virtually abandon judicial review . . . . The determination of reasonableness must take into account the extent to which the constitutional right—in this case the right to sue for a full recovery under Article I, section 11—is diminished and the extent

such an important right, Justice Stewart applied this heightened level of scrutiny in finding the challenged law unconstitutional.[135]

Based on the opinions of Justices Durham, Zimmerman, and Stewart, it appears that a majority of the court would apply some form of heightened scrutiny in an equal protection analysis involving a right protected under the Open Courts Provision. As Justice Durham summarized for the plurality, the traditional, almost total, deference afforded non-fundamental rights under federal equal protection analysis "is inappropriate when dealing with the fundamental principle of American law that victims of wrongful or negligent acts should be compensated to the extent that they have been harmed."[136] This Comment, therefore, will

---

> to which the burden imposed actually furthers the legislative goals, as well as the importance of those goals.

Id. (Stewart, J., concurring). Comparison of this language with the language of *Malan* suggests Stewart believes the level of scrutiny should be more searching than that employed in *Malan*. *Compare id.* at 373 (Stewart, J., concurring) ("[C]lassification must in fact reasonably and substantially further the legislative purpose.") *with Malan*, 693 P.2d at 670 ("[C]lassifications . . . must be based on differences that have a reasonable tendency to further the objectives of the statute.").

135. *Condemarin*, 775 P.2d at 372-74 (Stewart, J., concurring). Moreover, Justice Stewart would do so without the benefit of presuming constitutionality. *Id.* at 370. Unlike the *Berry* analysis, it is possible that a majority of the court would eliminate the presumption of constitutionality in an equal protection analysis involving an open-court-type right and conduct its own independent analysis. *See supra* note 103 (discussing presumption of constitutionality under *Berry* test and Justices Durham and Zimmerman's willingness to deny the benefit of presumption of constitutionality under the *Berry* test).

136. *Condemarin*, 775 P.2d at 354 (plurality opinion). Chief Justice Hall's dissent in *Condemarin*, in which Associate Chief Justice Howe concurred, is not necessarily to the contrary. Justice Hall applied a minimal level of scrutiny to the Act's recovery limitation and found it constitutional. *Id.* at 383-84, 387 (Hall, C.J., dissenting). His opinion recognized that the level of scrutiny varied depending on the interest involved and that the Utah Constitution may give rise to an analysis independent from federal precedent. *See id.* at 379-80 (Hall, C.J., dissenting). The Chief Justice nonetheless rejected the heightened *Malan* analysis, apparently because the right to a full recovery from the government was properly limited under the doctrine of sovereign immunity and was not protected by the Open Courts Provision. *See id.* at 383 (Hall, C.J., dissenting). In reaching his conclusion, he reasoned that at the time the Utah Constitution was adopted, plaintiffs had no right to recovery under the doctrine of sovereign immunity. *Id.* at 383-84 (Hall, C.J., dissenting). Thus, the Open Courts Provision offered no protection from limitations on legal recovery from governmental entities, and the right to full legal recovery from such entities was not fundamental. *Id.* (Hall, C.J., dissenting). Accordingly, Hall reasoned, the facts of the case did not trigger heightened scrutiny. *Id.* at 383-84 (Hall, C.J., dissenting). The language of the dissenting opinion thus leaves open the option that Chief Justice Hall and Associate Chief Justice Howe would find it appropriate to apply some form of heightened scrutiny in a case involving a right clearly

rely upon the heightened equal protection standard suggested by the *Condemarin* plurality to evaluate the Malpractice Act's treatment of minors.

4. Summary of Utah's Constitutional Analysis

As established by the review above, the Utah Supreme Court recently has shown a willingness to provide a meaningful constitutional review of legislation that infringes the right to recover for tortious injury. This is true whether the challenge is brought under the Utah Constitution's open courts or equal protection guarantees. Although the precise analytical method differs between the two guarantees, it appears that under either method a majority of the court will likely approach the issue of the Malpractice Act's treatment of minors in a way that carefully considers the reasonableness of the legislature's policy decision.

V. THE CONSTITUTIONALITY OF ELIMINATING TOLLING PRIVILEGES FOR MINORS BRINGING MEDICAL MALPRACTICE CLAIMS UNDER UTAH LAW

A. Utah's Open Courts Provision and the Berry *Test*

Although the issue is not entirely clear, it appears likely that a majority of the Utah Supreme Court would find the Malpractice Act's treatment of minors unconstitutional under the *Berry* test. As a threshold matter the court must first determine whether the right infringed is protected under the Open Courts Provision.[137]

---

protected under the Open Courts Provision.

It is also worth noting that in previous cases both Justices expressed concern about the constitutionality of eliminating tolling provisions for minors. *See* Myers v. McDonald, 635 P.2d 84, 88 (Utah 1981) (Howe, J., concurring) (indicating that legislature's failure to take into account "the legal disability of a child" may violate minors' due process, equal protection, and open courts provision rights); Scott v. School Bd., 568 P.2d 746, 748 (Utah 1977) (Justice Hall stated in dicta that to deprive minor of tolling provision included in statute at issue would be "a denial of due process and equal protection"). Consequently, Chief Justice Hall and Associate Chief Justice Howe may be willing to apply heightened scrutiny to the Malpractice Act's application of a shortened limitations period to minors because it infringes on a right that was clearly established at the time the Utah Constitution was adopted: a minor's right to personally bring an action upon attaining the age of majority.

137. *See supra* note 103 and accompanying text (discussing *Berry* test).

If the challenged legislation does infringe on such a right, the court must then determine whether a comparable substitute remedy has been provided.[138] If no substitute is provided, the challenged act is justifiable only if the court concludes that: (1) legislation eliminates a clear social or economic evil; and (2) the means chosen to eliminate the evil is not arbitrary or unreasonable.[139] This part of the test in essence balances the importance of the governmental interest at stake against the importance of the individual rights being diminished.[140] As summarized by Justice Durham:

> [T]he [*Berry*] Court examined the legitimacy of the legislative purpose and the extent to which said purpose was reasonably and substantially advanced by the means utilized and compared those "benefits" to the denial of rights protected by article I, section 11. . . . Legislative attempts to abrogate those rights should be closely examined by this Court and struck down when the disability they seek to impose on individual rights is too great to be justified by the benefits accomplished or when the legislation is simply an arbitrary and impermissible shifting of collective burdens to individual citizens.[141]

As illustrated below, the Malpractice Act's treatment of minors appears to be both an unjustified imposition on individual rights and an arbitrary shifting of collective burdens to a small group of unrepresented minors.

As an initial matter, a minor's right to bring an action in person upon attaining the age of majority is protected under the Open Courts Provision. First, the right of a person to bring a claim for negligent infliction of personal injury is clearly protected under the plain language of the constitutional provision.[142] Furthermore, a minor's right to bring a claim against another party for a negligent injury has been recognized since at least 1895.[143]

---

138. *See supra* note 103 and accompanying text (discussing *Berry* test).

139. *See supra* note 103 and accompanying text.

140. *See* Horton v. Goldminer's Daughter, 785 P.2d 1087, 1094 (Utah 1989).

141. *Condemarin*, 775 P.2d at 358 (plurality opinion); *see also Horton*, 785 P.2d at 1094 (characterizing open courts' analysis as balancing test).

142. The Open Courts Provision states: "All courts shall be open, and every person, for an injury done to him in his person . . . shall have remedy by due course of law . . . ." UTAH CONST. art. I, § 11.

143. *See* Chipman v. Union Pac. Ry., 41 P. 562, 562 (Utah 1895).

Moreover, since 1876, Utah law has prohibited minors from personally bringing an action in court, but has allowed them to bring a cause of action upon reaching the age of majority by tolling the applicable statute of limitation.[144] To eliminate the benefit of the tolling provision, therefore, is to completely abrogate a particular form of remedy—a personal action on the minor's own behalf—that has been available to minors for over 100 years.[145]

Thus, because the Malpractice Act impinges on a right protected by the Open Courts Provision, the *Berry* test mandates inquiry into whether the Malpractice Act provides a comparable substitute remedy. The Malpractice Act, however, provides no alternative remedy "substantially equal in value or other benefit to the remedy abrogated in providing essentially comparable substantive protection to one's person."[146] Arguably, a substitute is provided in place of a minor's right to bring an action upon attaining the age of majority because a parent or a guardian may bring the action for the minor.[147] The Utah Supreme Court has explicitly held, however, that an existing alternative remedy is not an adequate substitute.[148] Instead, the Legislature must offset the loss of a particular right by providing *quid pro quo* in the form of a substitute remedy,[149] something the Malpractice Act clearly fails to supply.[150]

---

144. *See supra* notes 2-7 and accompanying text (discussing early Utah disability and tolling provisions).

145. Indeed, the Malpractice Act's restriction is substantially greater than the limitation on damages provision struck down as unconstitutional in *Condemarin*. Whereas *Condemarin* involved a restriction on the amount of recoverable damages, the Malpractice Act's treatment of minors goes further by potentially barring minors from pursuing any remedy at all.

146. Berry v. Beech Aircraft Corp., 717 P.2d 670, 680 (Utah 1985).

147. *See* UTAH R. CIV. P. 17(b) (providing for minor's representation by guardian).

148. Sun Valley Water Beds v. Herm Hughes & Son, 782 P.2d 188, 192 (Utah 1989). According to the *Sun Valley* court, an existing alternative remedy is not sufficient to fulfill the requirement that the legislature offer a suitable substitute remedy: "[W]hen the legislature removes a particular right or remedy, it cannot simply rely on other preexisting rights or remedies to fill the void left behind . . . ." *Id.*

149. *Id.*

150. For similar reasons, the Legislature cannot seek recourse in the argument that in the event a parent or guardian fails to bring a cause of action within the specified period of time, the minor would have a substitute remedy against the parent or guardian for negligence. This argument fails because the Utah Supreme Court has expressly held that having a claim against a different party does not fulfill the substitute remedy requirement. *Id.*

Turning to *Berry*'s balancing test, the legislative restriction on a protected right is valid only if it is outweighed by an attempt to eliminate a "clear social or economic evil" and the means used is not "arbitrary or unreasonable."[151] On the one hand, there is the long standing and cautiously guarded right of a minor to seek remedies in court for personal injury,[152] a right clearly protected by Utah's Open Courts Provision.[153] On the other hand, there is the stated legislative purpose of protecting public access to health care.[154] It is unclear, however, whether growing malpractice litigation qualifies as a "clear social and economic evil" that must be eliminated in order to protect the public's access to health care.

Even if existing alternative remedies could be considered substitutes, they hardly rise to the level of "comparable substantive protection," when compared with a minor's ability to bring an action sometime after reaching the age of majority. Undoubtedly, parents and guardians will at times fail to adequately represent children's interests. A quick review of current headlines regarding growing incidents of child abuse provides the most obvious example of the fallibility of assuming that parents or guardians will always adequately represent minors' interests. Additionally, parents or guardians may fail to bring an action in a timely manner. *See, e.g.*, Scott v. School Bd., 568 P.2d 746, 746-47 (Utah 1977) (parent neglected to file notice within statutory period).

Furthermore, injured minors do not always have a parent or guardian to take action. *See, e.g.*, Gasparro v. Horner, 245 So. 2d 901, 905-06 (Fla. Dist. Ct. App. 1971) (holding that statute of limitations not tolled during orphan's minority). Consequently, the ability to rely on a parent or guardian is of lesser value than having the statute of limitation tolled until the disability of minority is removed.

Similarly, bringing an action against a parent for failure to institute an action is not a sufficient substitute. First, a minor's parents are under no duty to bring an action. *See Scott*, 568 P.2d at 747. Moreover, even if parents had a duty to bring an action, the intrafamily immunity doctrine may bar such a suit. This is unclear, however, as the doctrine is currently unsettled in Utah law. *See* Bishop v. Nielson, 632 P.2d 864, 865-68 (Utah 1981). Additionally, common sense dictates that a suit against one's parents would be a distinctly less favorable option than one against an insured physician or hospital, both in terms of securing sufficient compensation for the victim and preserving the family. *See* Mominee v. Scherbarth, 503 N.E.2d 717, 722 (Ohio 1986) ("Placing young adults in a dilemma in which they must choose between suing their parents or abandoning their claims has the practical effect of chilling their due process rights."). Lastly, recovering in a suit against a parent or guardian would present distinct difficulties; not only would the minor have to prove malpractice on the part of the health care provider, but also negligence on the part of the parent or guardian. *See id.* Consequently, an action against a negligent parent or guardian scarcely seems to be a comparable substitute under the *Berry* test.

151. Berry v. Beech Aircraft Corp., 717 P.2d 670, 680 (Utah 1985).

152. *See supra* notes 16-29 and accompanying text (discussing historical importance of safeguarding minors' legal rights).

153. *See supra* notes 142-45 and accompanying text (discussing applicability of Open Courts Provision to minors' rights).

154. *See supra* note 50 (quoting stated purposes of Malpractice Act).

The basic objective of the Malpractice Act is to "protect and insure the continued availability of health care services to the public."[155] This objective is undoubtedly a proper and legitimate legislative goal.[156] Moreover, there is no dispute that there are real and substantial concerns about public access to medical care.[157] Nor do most observers dispute that there are substantial problems with the current inefficiency of the tort system in compensating malpractice victims.[158]

The Malpractice Act is premised, however, on the belief that the growing numbers of malpractice claims pose a significant threat to public access to health care.[159] This belief can be questioned legitimately. The Malpractice Act's findings conclude that increasing claims and settlements impact availability in four ways: (1) health care cost is increased because providers pass the cost of insurance to the patient; (2) health care cost is also increased when the provider practices defensive medicine in response to a perception that a patient is a potential adversary in a lawsuit; (3) availability of health care services is adversely affected because malpractice insurance is sometimes unavailable; and (4) availability of health care services diminishes because providers are discouraged from providing services due to the high cost of malpractice insurance.[160] Each of these assumptions are

---

155. Allen v. Intermountain Health Care, Inc., 635 P.2d 30, 32 (Utah 1981); *see also* UTAH CODE ANN. § 78-14-2 (1992) (legislative statement of purpose of Malpractice Act).

156. Utah, like all other states, is clearly able to enact laws and take measures necessary to promote and preserve the public health. *See* Barsky v. Board of Regents, 347 U.S. 442, 449 (1954).

157. *See, e.g.*, UTAH HEALTH CARE ACCESS STEERING COMM., RECOMMENDATIONS FOR A UTAH HEALTH ACCESS PROGRAM FOR THE STATE OF UTAH 10-16 (1991) (describing national and local problems of access to health care for uninsured people); Health Care Falls Short Across U.S., SALT LAKE TRIB., Feb. 18, 1992, at A8 (estimating 43 million people nationally have poor access to medical care and Utah ranks 33rd in poor access with 11.4% of Utah's population underserved by medical system).

158. *See, e.g.*, WEILER, *supra* note 36, at 52-54, 159-61 (noting inefficient delays and administrative costs current tort system imposes in compensating malpractice victims). Criticisms of the current compensation system include its failure to identify and redress a high number of valid claims, the large percentage of insurance funds that go to legal fees and other administrative costs of the system, the lengthy delays in compensating victims, and its questionable value in deterring negligent medical care. *Id.*

159. *See supra* note 50 (quoting Malpractice Act's legislative findings).

160. *See* UTAH CODE ANN. § 78-14-2 (1992). The Legislature's findings comport with three basic components of the health care costs associated with medical malpractice: first, medical malpractice insurance costs and availability; second, defensive medicine costs (medical care provided more for a potential legal defense than for value in health

questionable, however, particularly when viewed collectively as a basis for shortening the statute of limitation.

First, increased insurance costs are undoubtedly passed on to the consumer. It is questionable, however, if the impact of those costs is significant enough to warrant substantial restriction on the right to recover for malpractice. Numerous studies show that increases in malpractice insurance costs comprise but a minor fraction of the growth in the overall cost of health care; estimates generally place the cost of malpractice insurance at about one percent of the nation's health care bill over the last three decades.[161] In other words, to the extent that burgeoning health care costs impede public access to health care, the impact of malpractice insurance costs is a relatively minor concern.[162]

---

care); and finally the intangible costs resulting from physicians being unwilling or unable to practice in certain specialties or locations due to malpractice concerns. *See* SLOAN et al., *supra* note 36, at 11-12; HEW REPORT, *supra* note 35, at 12-16.

161. *See, e.g.*, WEILER, *supra* note 36, at 4 (malpractice insurance as a share of nation's total health bill "rose from just under 0.5 percent [in 1960] to a little over 1.0 percent [in 1988]"); *cf.* DANZON, *supra* note 41, at 186-87 (cost of malpractice insurance just over one percent of medical care expenditures in 1975 and less than one percent in 1982); Sylvia A. Law, *A Consumer Perspective on Medical Malpractice*, LAW & CONTEMP. PROBS., Spring 1986, at 305, 308 (malpractice premiums make up about one percent of total national cost of health care); Sarah Glazer, *Whatever Happened to the Malpractice Insurance Crisis?*, WASH. POST, July 9, 1991, Magazine, at 10, 11 (insurance costs barely one percent of $600 billion annual health care bill and about three percent including American Medical Association's estimated $15 billion in defensive medicine costs).

162. *Cf.* Mominee v. Scherbarth, 503 N.E.2d 717, 724 (Ohio 1986) (Celebrezze, C.J., concurring) ("[T]he assurance of the availability of health care in Ohio is an institutional problem of sweeping dimension that will not be solved merely by reducing medical malpractice insurance premiums."). Indeed, according to one observer, "'[t]he notion that you're going to solve health care costs through reforming the malpractice system is an illusion.'" Glazer, *supra* note 161, at 11 (quoting Peter D. Jacobson, Senior Behavioral Scientist, Rand Corporation). *But cf.* DANZON, *supra* note 41, at 187 ("Nevertheless, the burden of litigation costs as a share of the total is high enough to warrant serious search for reform."). In addition, to the extent that injuries suffered as a result of malpractice are not compensated by malpractice insurance, some of the expense will obviously be borne by private health insurers, where it will similarly impact the cost of access to health care.

Notably, increasing malpractice insurance costs have not created an issue of affordability for most physicians; according to a number of sources, on average physicians have been able to preserve their real net income position throughout the 1980s. WEILER, *supra* note 36, at 4-5; *see also* DANZON, *supra* note 41, at 187 (percentage of insurance expense to physician income remained comparable from 1970 to 1981); Law, *supra* note 161, at 308 (proportion of gross income to malpractice premiums paid by average practitioner decreased from 4.40% in 1976 to 3.69% in 1983).

Second, the notion that malpractice litigation promotes wasteful defensive medicine is well accepted.[163] There is controversy, however, over the extent of defensive medicine practices and the extent to which they are actually a negative phenomenon.[164] More significantly, it is questionable whether defensive medicine costs would be significantly mitigated by shortening limitation periods. While such procedural reform may reduce the number of claims, it does little to change the underlying system or eliminate providers' fears of litigation.[165] Consequently, although the cost of defensive medicine may be significant, it is unclear whether limiting access to the courts will significantly reduce those costs.

Third, it appears that insurance availability may have been a concern in certain states during the 1970s. In Utah, however, it is unclear whether there was ever a period of unavailability.[166] In any case, availability problems largely have been resolved by the formation of physician-sponsored insurance companies.[167]

Fourth, the concern that physicians may voluntarily withdraw medical care in response to malpractice costs and concerns is potentially more troubling than the impact of increased health care costs induced by malpractice litigation. For example, growing insurance costs and the risk of being accused of malpractice may lead a physician to retire early or eliminate a high-risk type of service.[168] Some physician surveys indicate a significant reduction in services due to malpractice insurance concerns or fears of

---

163. WEILER, *supra* note 36, at 85.

164. *See id.* at 85-89; DANZON, *supra* note 41, at 146-49; SLOAN et al., *supra* note 36, at 11.

165. *See* Randall R. Bovbjerg, *Problems and Solutions in Medical Malpractice*, in THE LIABILITY MAZE: THE IMPACT OF LIABILITY LAW ON SAFETY AND INNOVATION 274, 285 (Peter W. Huber & Robert E. Litan eds., 1991); *see also* WEILER, *supra* note 36, at 37 (procedural reforms reducing access to courts affect types of claims brought "but they will not generally have a large impact on the amount of litigation doctors have to face or the premiums they have to pay").

166. *See* OFFICE OF LEGISLATIVE RESEARCH, STATE OF UTAH, MEDICAL MALPRACTICE INSURANCE PROBLEMS: REPORT TO THE 41ST LEGISLATURE 6-7 (1976) (noting that Utah has not "yet been without medical malpractice insurance coverage").

167. *See* SLOAN et al., *supra* note 36, at 207. This is clearly illustrated in Utah. For example, the physician-sponsored Utah Medical Insurance Association currently provides about 71% of the coverage in the state, while 30 other insurance companies cover the rest. UTAH INS. DEP'T, REPORT TO THE GOVERNOR: BUSINESS OF 1990, at 58 (1991).

168. Stephan Zuckerman et al., *Information on Malpractice: A Review of Empirical Research on Major Policy Issues*, LAW & CONTEMP. PROBS., Spring 1986, at 85, 106, 109-110 (discussing obstetricians).

litigation.[169] Data supporting this concept, however, is largely anecdotal.[170] More significantly, it is unclear whether tort reforms limiting access to courts will limit liability sufficiently to significantly induce practitioners to cease curtailing their services.[171]

Reviewing the assumptions underlying the expressed legislative purpose of the Malpractice Act illustrates that it is not entirely clear whether increases in malpractice litigation significantly hinder public access to health care.[172] Highlighting the

---

169. *Id.*; *see also* 1 COMMITTEE TO STUDY MEDICAL PROFESSIONAL LIABILITY AND THE DELIVERY OF OBSTETRICAL CARE, INSTITUTE OF MEDICINE, MEDICAL PROFESSIONAL LIABILITY AND THE DELIVERY OF OBSTETRICAL CARE 35-52, 150-51 (1989) [hereinafter INSTITUTE OF MEDICINE] (concluding availability of obstetrical care is reduced due to malpractice costs and litigation, particularly in rural areas).

170. SLOAN et al., *supra* note 36, at 11. According to this source, surveys of obstetricians indicating substantial reductions in access must be viewed skeptically because the data is "self-reported in an atmosphere of crisis and political mobilization." *Id.* at 11-12. "Moreover, although a physician who curtails his or her practice clearly loses access to certain patients, it is less clear that his or her former patients lose access to care. Virtually nothing reliable is known about impacts on patient access to care." *Id.* at 12; *see also* Zuckerman et al., *supra* note 168, at 109-10 ("Unless a survey of patients shows that certain types of high-risk care is becoming increasingly difficult to obtain and that this difficulty is correlated with aspects of malpractice activity, however, the provider-based results need to be viewed cautiously."). Although there are unsubstantiated anecdotal references to reduced access to care in rural Utah, *see, e.g.*, UTAH HEALTH CARE ACCESS STEERING COMM., *supra* note 157, at 25 (asserting that in rural Utah a decreasing number of physicians are willing to deliver babies), there does not appear to be any evidence documenting such a phenomenon in Utah or that it has caused substantial malpractice concerns.

171. *See supra* note 162 (discussing inadequacy of procedural reforms in resolving problems caused by malpractice litigation).

172. Indeed, since the beginning of the so-called malpractice crisis, courts and commentators have disputed whether a crisis in fact exists. *See, e.g.*, Jones v. State Bd. of Medicine, 555 P.2d 399, 411-16 (Idaho 1976) (court unwilling to rely on legislative declarations and remanding case for lack of factual record establishing existence of malpractice crisis in state or nation); Boucher v. Sayeed, 459 A.2d 87, 92-94 (R.I. 1983) (questioning continued existence of crisis); Cunningham & Lane, *supra* note 36, at 114-21 (questioning existence of crisis). Instead, according to one observer, "[r]egardless of whether the underlying circumstances were truly severe, the perceptions of a panicked public as well as ferocious lobbying by the medical profession and insurance industry generated intense pressure on state legislatures to enact remedial legislation." Howard A. Learner, Note, *Restrictive Medical Malpractice Compensation Schemes: A Constitutional "Quid Pro Quo" Analysis to Safeguard Individual Liberties*, 18 HARV. J. ON LEGIS. 143, 144 (1981) (footnote omitted).

More specifically, the same criticism can be aimed at Utah. Based on malpractice insurance studies prepared for the Utah Legislature, it is not clear at the time the Malpractice Act was enacted whether Utah was actually suffering from a malpractice insurance crisis. *See* MALPRACTICE INS. ADVISORY COMM., A REPORT ON MEDICAL MALPRACTICE TO THE ECONOMIC DEVELOPMENT COMMITTEE OF THE UTAH LEGISLATIVE

weaknesses of the assumption that malpractice litigation substantially adversely affects public access to health care, however, is not intended to minimize the negative impact that does occur. Instead, the purpose of this review is to demonstrate that the Malpractice Act is not based on a clear and well-documented relationship between malpractice litigation and health care availability.[173]

Therefore, it is not entirely clear whether the alleged crisis in malpractice claims rises to a level that the Utah Supreme Court would consider a "clear social or economic evil" outweighing the harsh treatment of minors who may fail to bring timely medical malpractice claims.[174]

Ironically, the large number of malpractice victims that receive no compensation under the current tort system *is* a clear social or economic evil that needs to be addressed in this area. Several sources indicate that the number of persons injured by

---

COUNCIL 2-6 (1970) (extent of national insurance problems present in Utah unclear as little or no statistical information on Utah situation available); OFFICE OF LEGISLATIVE RESEARCH, *supra* note 166, at 6-7 (Utah experiencing same trends in claims as other states, but problem is not of same magnitude).

Significantly, neither of the two studies prepared for the Utah Legislature on medical malpractice evaluated the question of whether, or to what extent, increases in malpractice insurance might affect the accessibility of health care. *See* MALPRACTICE INS. ADVISORY COMM., *supra*; OFFICE OF LEGISLATIVE RESEARCH, *supra* note 166. The debate over whether there is a malpractice insurance crisis in Utah continues today. *See, e.g.*, Access To Health Care Task Force, Minutes from Task Force July 30 Meeting, at 3-5 (1991) (witnesses before task force disputing extent of malpractice crisis in Utah); *cf.* Glazer, supra note 161, at 10-13 (discussing current debate over continued existence of a national malpractice insurance crisis).

173. *See generally* Zuckerman et al., *supra* note 168, at 110-11 (discussing general lack of empirical evidence demonstrating relationship between claims, lawsuits, and premiums on the practice of medicine, and noting that policymakers are making decisions on malpractice reform issues based on little hard evidence).

174. The uncertainty over whether malpractice litigation is an "evil to be eliminated" also stems from the fact that no opinion applying the *Berry* test has clearly delineated this portion of the test. Thus, it is unclear how rigorous the standard is. However, there has been only one case in which the court has concluded that the challenged legislation failed to address any "clear social or economic evil." Horton v. Goldminer's Daughter, 785 P.2d 1087, 1094 (Utah 1989). In *Horton*, the court held unconstitutional a statute of repose limiting actions against builders and architects in part because the apparent objective—providing relief to builders and architects from perpetual risk of liability for latent defects—was not sufficient. *Id.* at 1094-96.

Most of the cases applying the *Berry* test instead focus on the unreasonable or arbitrary nature of means used to achieve the purported legislative objectives. *See* Sun Valley Water Beds v. Herm Hughes & Son, 782 P.2d 188, 192-94 (Utah 1989); Berry v. Beech Aircraft Corp., 717 P.2d 670, 680-83 (Utah 1985). Nevertheless, it seems appropriate to at least consider whether malpractice litigation is the evil that the Legislature concludes it to be.

malpractice is far larger than the number filing claims.[175] One recent study, for example, indicated that only about two percent of patients injured by medical malpractice ever file claims.[176] This suggests that if there is a fundamental cause of the malpractice insurance problem it is the amount of malpractice occurring in the country.[177] Yet, to the extent it reduces access to the legal system, the Malpractice Act makes it harder for victims to legitimately seek compensation. From this perspective, it seems somewhat ironic that the growing number of persons willing to bring claims for injury—a constitutionally protected right—should be viewed as an evil to be eliminated. Rather, the proper focus of the Legislature should be on the efficient compensation of all malpractice victims, including those that for some reason are left out of the current system.[178]

Nevertheless, even assuming the legislative purpose is sufficient to meet *Berry*'s clear social and economic evil requirement, the Malpractice Act's total abrogation of minors' rights fails the final portion of the *Berry* test because it is an unreasonable and arbitrary means of addressing the problem. First, the method is arbitrary because there does not appear to be any significant statistical evidence demonstrating that the elimination of a "long

---

175. *See, e.g.*, DANZON, *supra* note 41, at 18-25 (estimating between one in five and one in ten malpractice injuries result in claims); Jenkins & Schweinfurth, *supra* note 38, at 851 n.127 (citing numerous sources finding many more injuries occur than claims indicate).

176. *See* A. Russell et al., *Relation Between Malpractice Claims and Adverse Events Due to Negligence*, 325 NEW ENG. J. MED. 245, 247 (1991).

177. *See* Jenkins & Schweinfurth, *supra* note 38, at 851 (among many factors contributing to malpractice insurance crisis, chief cause is high incidence of actual medical malpractice occurring throughout country).

178. According to one observer:

> I think my medical friends are making a grave mistake when they propose reforms that uniformly make it harder for people who get hurt to get paid. As badly as physicians are getting hurt, they are not getting hurt like the people who get hurt in the course of health care. Not one of their reforms . . . helps an injured person get paid. Every one of them either makes it harder for him to get paid or pays him less.
>
> . . . .
>
> . . . We ought not to be thinking about just restricting the right of the injured patient. We ought not to be going to the extreme that pays every patient who gets "adversely affected" by health care. We ought to be trying to find a middle ground that tries to pay people without having to litigate endlessly what happened in the course of health care.

Jeffrey O'Connell, *Overview of the Liability System in* MEDICAL MALPRACTICE 4, 6 (Duncan Yaggy & Patricia Hodgson eds., 1987).

tail" problem with minors will have a significant impact on medical malpractice insurance rates.[179] Indeed, the scarce statistical information regarding medical malpractice claims by minors in Utah indicates the minimal impact of these claims. For example, one study found that in Utah, minors made only 10.9% of all claims (not lawsuits) during a 3.5 year period.[180] Furthermore, several studies from other states indicate that greater than ninety percent of all claims are brought within five years of the malpractice incident.[181] Combining these figures roughly suggests that less than one percent of claims in a given year involve a minor bringing a claim after five years.[182] The potential impact of minors filing delayed claims becomes even more attenuated considering studies that indicate about half of all claims are

---

179. Neither of two Utah legislative reports on the malpractice insurance problem provided any data indicating the number of claims by minors or indicated the existence of a "long tail" problem in the state. *See* MALPRACTICE INS. ADVISORY COMM., *supra* note 172; OFFICE OF LEGISLATIVE RESEARCH, *supra* note 166. Interestingly, neither did they discuss or demonstrate a significant relationship between increasing malpractice insurance costs and decreasing accessibility to health care. *See* MALPRACTICE INS. ADVISORY COMM., *supra* note 172; OFFICE OF LEGISLATIVE RESEARCH, *supra* note 166.

At least one source doubts the existence of the "long tail" problem. COMMISSION ON MEDICAL PROFESSIONAL LIABILITY, AM. BAR ASS'N, INTERIM REPORT OF THE COMMISSION ON MEDICAL PROFESSIONAL LIABILITY App. E, at 60, 62-63 (1976) [hereinafter ABA INTERIM REPORT] ("The data reviewed by this commission has demonstrated that the terror of the 'long tail' for medical insurance liability is substantially a myth.").

180. *See* NATIONAL ASS'N OF INS. COMM'N, MALPRACTICE CLAIMS: FINAL COMPILATION, MEDICAL MALPRACTICE CLOSED CLAIMS 1975-1978, at 115 (1979) [hereinafter NAIC REPORT]. According to the records of insurers writing premiums of $1,000,000 or more in any year since 1970, of the 237 claims made in Utah during a 3 1/2 year period, 26 were made by minors, for an average of about 7.5 claims per year. *Id.* Of 61,624 malpractice claims brought nationally against such insurers during the same period, 12.5% were made by minors. *Id.*; *cf.* HEW REPORT, *supra* note 35, app. at 12 (14.4% of all claims filed over a 10 year span ending in 1970 were made by or on behalf of individuals under the age of 20).

181. *See* NAIC REPORT, *supra* note 180, at 115 (12 month average time from incident until report for claims paid for minors in Utah during a three and a half year period; 25 month national average); ABA INTERIM REPORT, *supra* note 179, at 19 n.8 (approximately 94% of all claims reported within three years of injury); Kenneth S. Abraham, *Medical Malpractice Reform: A Preliminary Analysis*, 36 MD. L. REV. 489, 503 (1977) (citing study showing 98.1% of medical malpractice incidents reported within five years of occurrence); DANZON, *supra* note 41, at 182 (citing study showing 98.2% of incidents reported within five years of injury). *But see* DANZON, *supra* note 41, at 183 (study showing 63-75% of claims made within three years of injury).

182. Applying these figures to the 26 claims made by Utah minors as reported in the closed claims study by the NAIC Report roughly suggests that on average less than one minor per year would bring a claim later than five years after the incident. *See* NAIC REPORT, *supra* note 180, at 115.

settled without any payment.[183] As such, the meager statistics suggest that the "long tail" problem in Utah is negligible.[184]

Moreover, the legislature's conclusion that a minor's interest can adequately be represented by parents or guardians is questionable. Although many minors will be adequately represented, inevitably some minors will be completely denied the right to bring a cause of action, through no fault of their own, because a parent or guardian (assuming one exists) fails to take action. As the Arizona Supreme Court observed:

> While the vast majority of claims on behalf of injured minors will still be brought within a relatively short time after the injury occurs, this all depends upon good fortune; the minor himself is helpless, particularly when under ten years of age. The minor possesses a right guaranteed by the constitution, but cannot assert it unless someone else, over whom he has no control, learns about it, understands it, is aware of the need to take prompt action, and in fact takes such action.
>
> . . . The statute makes no exceptions for children who have unconcerned parents, children in foster care, or those in institutions; it applies alike to children who are precocious and those who are retarded, those who are normal and those who are brain injured. It applies to those with guardians and those without.[185]

Viewed in this light, the Malpractice Act is unreasonable because it fails to distinguish between minors who have a parent or guardian able or available to bring a claim and those who do

---

183. *See, e.g.*, WEILER, *supra* note 36, at 42 (90% of claims settled out of court and half of those closed without payment); NAIC REPORT, *supra* note 180, at 115 (during three and one-half year period 64% of claims closed without payment in Utah, 58% nationally); U.S. GEN. ACCOUNTING OFFICE, MEDICAL MALPRACTICE: CHARACTERISTICS OF CLAIMS CLOSED IN 1984, at 2 (1987) (43% of claims studies resulted in payment).

184. *Cf.* Berry v. Beech Aircraft Corp., 717 P.2d 670, 682 (Utah 1985) (because only 2.6% of product-related injuries occurred after six years from date of purchase, number of cases arising after statute of limitations period is insignificant); Mominee v. Scherbarth, 503 N.E.2d 717, 721 (Ohio 1986) (questioning the real significance of the relation between minors claims and reducing medical malpractice claims); Jenkins & Schweinfurth, *supra* note 38, at 960-62 (questioning whether elimination of "long tail" effect for minors will decrease insurance costs).

185. Barrio v. San Manuel Div. Hosp., 692 P.2d 280, 286 (Ariz. 1984); *see also* Sax v. Votteler, 648 S.W.2d 661, 667 (Tex. 1983) (unreasonable to rely on notion that parents or guardians will always protect interest of minors).

not.[186] Nor does the Malpractice Act distinguish infants who cannot effectively communicate physical problems resulting from an injury from older children who can.[187]

Finally, Utah's stance on treatment of minors in medical malpractice actions is unreasonable when compared with other states. The great majority of states provide tolling protection either by retaining a reduced tolling period for minors,[188] allowing

---

186. *Cf. Barrio*, 692 P.2d at 286 (noting statute reducing infant tolling provision makes no exceptions for children not adequately represented).

The failure of a parent to bring an action on behalf of a child may not necessarily be due to neglect or ignorance. In some cases, a parent's failure may be due to the difficulty the parent has coming to grips with the fact that their child may be seriously injured, leading the parent to deny the extent, or existence, of an injury. For example, when discussing the evaluation of brain-damaged children, one commentator noted "[p]arents . . . have built-in defenses, some more than others. Some can be told again and again and again about the nature of a baby's problems—and they still will not listen." Ellison, *supra* note 47, § 24.02.

187. *See supra* notes 43-48 and accompanying text (describing how most malpractice reforms distinguish between very young plaintiffs and older minors); *cf. Berry*, 717 P.2d at 681 (holding products liability statute of repose arbitrary because it applies to all kinds of products irrespective of unique product features).

188. *See* ALA. CODE § 6-5-482(b) (1975) (minors under age four have until age eight to file action); CAL. CIV. PROC. CODE § 340.5 (West 1982) (minors under age six have three years or until age eight to file action, whichever is longer); COLO. REV. STAT. § 13-80-102.5(3)(d)(I) (Supp. 1991) (minors under age six have until age eight to file action); DEL. CODE ANN. tit. 18, § 6856(2) (1989) (minors under age six have two years or until age six whichever is longer); GA. CODE ANN. § 9-3-73(b) (Supp. 1992) (minors under age five have until age seven to file an action); HAW. REV. STAT. § 657-7.3 (1988) (minors under age six have six years or until age ten to file action whichever is longer); IDAHO CODE § 5-230 (1991) (tolling limitation period for six years); ILL. ANN. STAT. ch. 110, para. 13-212(b) (Smith-Hurd Supp. 1991) (tolling limitation period for eight years or until age 22 whichever is longer); IND. CODE ANN. § 16-9.5-3-1(a) (Burns 1990) (minors under age six have until age eight to file an action); KAN. STAT. ANN. § 60-515(a) (1983) (minors have extended eight year statute of repose); ME. REV. STAT. ANN. tit. 24, § 2902 (1964) (minors have extended six year statute of limitation or three years after attaining age of majority, whichever is shorter); MD. CODE ANN., CTS. & JUD. PROC. § 5-109(b)-(c) (1989) (minors may toll to age eleven; minor with injury to reproductive system may toll to age sixteen); MASS. ANN. LAWS ch. 231, § 60D (Law. Co-op. Supp. 1991) (minors under age six have seven years or until age nine to file action); MICH. STAT. ANN. § 27A.5851(7) (Callaghan Supp. 1991-92) (minors under age thirteen have until age fifteen to file action); MINN. STAT. ANN. § 541.15(a)-(b) (West 1988) (tolling limitation period for seven years or one year after attaining age of majority whichever is shorter); MISS. CODE ANN. § 15-1-36(2)-(3) (Supp. 1991) (limitation period tolled for minors without parent or guardian, otherwise minors under age six have until age eight to file action); MONT. CODE ANN. § 27-2-205(2) (1991) (limitation period tolled for minor not residing with parent or guardian, otherwise minors under age four have until age eight to file action); N.M. STAT. ANN. § 41-5-13 (Michie 1989) (minors under age six have until age nine to file action); N.Y. CIV. PRAC. L. & R. § 208 (McKinney 1990) (tolling limitation period for up to ten years); N.C. GEN. STAT. § 1-17 (Supp. 1991) (minors must file an action by age nineteen); N.D. CENT. CODE § 28-01-25 (1991) (tolling limitation period for up to 12

tolling for minors without a parent or guardian to represent them,[189] or allowing tolling for certain types of malpractice injury.[190] Utah, on the other hand, is one of only two states to abandon its historical approach of providing tolling protection in favor of the harsh position that both minors and adults bringing medical malpractice claims must meet the same restrictive statute of limitation.[191] Indeed, even the American Medical Association,

---

years); S.C. CODE ANN. § 15-3-545(D) (Law. Co-op. Supp. 1991) (tolling limitation period for seven years or one year after attaining age of majority whichever is shorter); TEX. REV. CIV. STAT. ANN. art. 4590i, § 10.01 (West Supp. 1992) (minors under age twelve have until age fourteen to file action); W. VA. CODE § 55-7B-4(b) (Supp. 1991) (minors under age ten have until age twelve to file action); WIS. STAT. ANN. § 893.56 (West 1983) (regular limitation period or by age ten whichever is greater); WYO. STAT. § 1-3-107 (1988) (minors have two years or by age eight to file action whichever is longer).

189. *See* ARK. CODE ANN. § 16-114-203(d) (Michie Supp. 1991) (for obstetrical care injuries, minor can toll to age nine or two years after minor acquires parent or guardian); MISS. CODE ANN. § 15-1-36(2)-(3) (Supp. 1991) (limitation period tolled for minors without parent or guardian, otherwise minors under age six have until age eight to file action); MONT. CODE ANN. § 27-2-205(2) (1991) (limitation period tolled for minors not residing with parent or guardian, otherwise minors under age four have until age eight to file action).

190. *See* ARK. CODE ANN. § 16-114-203(d) (Michie Supp. 1991) (for obstetrical care injuries, minor can toll to age nine or two years after minor acquires parent or guardian); MD. CODE ANN., CTS. & JUD. PROC. § 5-109(b)-(c) (1989) (minors may toll to age eleven; minor with injury to reproductive system may toll to age sixteen); NEV. REV. STAT. ANN. § 41A.097(3) (Michie Supp. 1991) (minor with brain damage or birth defect can toll to age ten; minor can toll until two years after minor attains age of majority for sterility caused by malpractice); WASH. REV. CODE ANN. § 4.16.350(3) (Supp. 1992) (minor can toll for one year after injury).

191. Oregon is the only other state to historically provide tolling protection for minors that takes the extreme position of eliminating all tolling protection in medical malpractice actions. *See* OR. REV. STAT. § 12.110(4), 12.160 (1991); *cf.* Fry v. Willamalane Park & Recreation Dist., 481 P.2d 648, 651 (Or. Ct. App. 1971) (applying former provision tolling limitation period for minors). Oregon's provision has withstood constitutional challenge. *See* Jones v. Salem Hosp., 762 P.2d 303, 308-09 (Or. Ct. App. 1988), *cert. denied*, 770 P.2d 595 (Or. 1989). Significantly, however, the Oregon court reviewed the provision only under a "rational basis" standard. *Id.* at 309.

Three other states, Connecticut, Florida, and Louisiana, also provide no tolling protection for minors in medical malpractice actions. *See* Lopez v. United Nurseries, Inc., 490 A.2d 1027, 1029 (Conn. App. Ct. 1985); University of Miami v. Bogorff, 583 So. 2d 1000, 1002 (Fla. 1991); Milton v. Merrell-Dow Pharmaceuticals, Inc., 589 So. 2d 1198, 1201 (La. Ct. App. 1991). These states are distinctly unlike Utah, however, as they historically have *not* provided general statutory tolling protection for minors bringing personal injury actions. *See, e.g.*, Lametta v. Connecticut Light & Power Co., 92 A.2d 731, 732-33 (Conn. 1952) (holding no tolling exception for disability of minority when bringing personal injury action); Gasparro v. Horner, 245 So. 2d 901, 905-06 (Fla. Dist. Ct. App.) (same), *cert. denied*, 249 So. 2d 689 (Fla. 1971); Steel v. Aetna Life & Casualty, 304 So. 2d 861, 863-64 (La. Ct. App. 1974) (applying standard limitation period to minor); *see also* 4 AM. JUR. TRIALS *Statutes of Limitation* app. at 602-03 (1966) (Connecticut, Florida, and Louisiana only three states in nation without a specific

the foremost advocate of reform in this area, recommends that a minimum tolling period run until age six or eight.[192]

In sum, it appears that under the Utah Supreme Court's *Berry* test, the Malpractice Act's abrogation of the tolling provision for minors would be unconstitutional. The Malpractice Act infringes on a right protected by the Open Courts Provision, yet fails to provide an acceptable alternative. Furthermore, the Legislature's choice to protect access to health care by restricting minors' rights is arbitrary and unreasonable. Even assuming malpractice insurance costs have a substantial effect on health care access, the few available statistics suggest the Malpractice Act's limitation statute for minors results in only minor, and possibly negligible, malpractice insurance savings.[193] Consequently, the Utah Supreme Court would likely find the Malpractice Act places an unconstitutional burden on minors who, through no fault of their own, do not have a parent or guardian willing or capable to bring an action on their behalf.

B. Utah's Equal Protection Analysis

As previously established, an equal protection analysis applied to this issue would likely involve a level of scrutiny greater than minimal rationality.[194] Under this test, the court must first determine that the classification used is reasonable and not arbitrary.[195] Next, the court must determine that the statute in fact reasonably and substantially furthers the legislative purpose.[196] As with the *Berry* test, this requires balancing: "The determination of reasonableness must take into account the extent to which the constitutional right . . . is diminished and the extent

---

disability tolling statute).

192. *See* AMA TASK FORCE, *supra* note 45, at 14. The original proposal suggested by the Social Services Study Committee to the Utah Legislature in 1976 suggested a tolling period for minors until the age of six. OFFICE OF LEGISLATIVE RESEARCH, *supra* note 166, at 26.

193. *Cf.* Strahler v. St. Luke's Hosp. 706 S.W.2d 7, 11 (Mo. 1986) (en banc) ("[W]e think the [shortened statute of limitation for minors] employed by the legislature to battle . . . escalating economic and social costs connected with medical malpractice litigation exacts far too high a price from minor plaintiffs . . . .").

194. *See supra* notes 106-36 and accompanying text (discussing appropriate level of review for Malpractice Act).

195. *See supra* notes 106-36 and accompanying text.

196. *See supra* notes 106-36 and accompanying text.

to which the burden imposed actually furthers the legislative goals, as well as the importance of those goals."[197] Thus, although the equal protection test is framed differently, the issue is resolved by considering many of the same factors considered under the *Berry* test.

Arguably, the classifications established by the Malpractice Act's limitation period are reasonable. Assuming restriction of malpractice claims is necessary to improve health care accessibility, the Malpractice Act seeks to restrict equally all malpractice claimants by applying the same shortened limitation period.[198] Similarly, assuming the "long tail" problem impacts health care availability, the Malpractice Act treats minors bringing malpractice actions in a class distinct from minors bringing other types of claims. In rebuttal, one might argue that these classifications are actually arbitrary given the uncertain validity of the Malpractice Act's underlying assumptions.[199] This argument is perhaps better made by focusing on whether the Malpractice Act's treatment of minors reasonably and substantially furthers the legislative purposes underlying the Malpractice Act. In this case, the Malpractice Act apparently does neither.

First, the near total lack of documented evidence of a "long tail" problem makes it questionable whether eliminating the tolling provision for minors substantially furthers the goal of reducing medical malpractice premiums. As discussed in the *Berry* analysis above, the meager existent evidence suggests that the number of minors likely to bring a delayed claim is negligible.[200] Indeed, it appears that on average less than one-half of one percent of all Utah Malpractice claimants are minors bringing claims more than five years after their injury.[201] Moreover, some observers conclude that medical malpractice litigation reforms generally have only a modest or negligible affect on limiting the

197. Condemarin v. University Hosp., 775 P.2d 348, 373 (Utah 1989) (Stewart, J., concurring).

198. *Cf.* Malan v. Lewis, 693 P.2d 661, 670 (Utah 1984) ("First, a law must apply equally to all persons within a class.").

199. *See supra* notes 151-74 and accompanying text (questioning relationship between restricting access to courts and improving public accessibility to health care).

200. *See supra* notes 179-84 and accompanying text (reviewing statistics on "long tail" problem with minors).

201. *See supra* notes 179-84 and accompanying text.

adverse effects of growing malpractice litigation claims.[202] Consequently, it is difficult to see how the Act's treatment of minors substantially furthers the goal of reducing medical malpractice premiums and the attendant adverse effects on health care availability.[203]

Second, the Malpractice Act's elimination of the tolling provision is an unreasonable method for fulfilling the Malpractice Act's legislative purpose. As established above, the Malpractice Act places a substantial burden on minors by completely abrogating their well-established right to personally bring an action. Thus, it seeks to treat minors as adults. It does so, however, despite minor's lack of legal standing and regardless of the unique conditions facing minors injured by malpractice. The Malpractice Act eliminates tolling regardless of whether minors have a parent or guardian to assert their claims for them, and despite the difficulties in discovering and evaluating injuries in young minors. In other words, the Malpractice Act's treatment of minors violates the general principle of equal protection, that "persons in different circumstances should not be treated as if their circumstances were the same."[204] Accordingly, the Malpractice Act's treatment of minors appears to be an unconstitutional means of minimizing medical malpractice insurance costs.

---

202. *See, e.g.*, INSTITUTE OF MEDICINE, *supra* note 169, at 127-31 ("[E]ven in cases where tort reforms have achieved the limited objective of reducing the size of awards, they have not lessened the tort system's negative impact on the delivery of obstetrical care nor have they increased providers' confidence in the system."); WEILER, *supra* note 36, at 37 (reforms limiting access to courts "are more likely to affect the timing and tactics employed by the parties in tort litigation, thereby influencing the types of medical accidents that will be channeled into the tort process; but they will not generally have a large impact on the amount of litigation doctors have to face or the premiums they have to pay").

203. Indeed, as one commentator noted, many of the malpractice reforms have been passed in a virtual vacuum of data substantiating the effect of tort reforms on insurance premiums: "The problem is simply that there is no evidence on which to base an evaluation of [insurance] claims, and what attempts lawmakers *are* making to address the problem are being made in the dark." Eliot M. Blake, Comment, *Rumors of Crisis: Considering the Insurance Crisis and Tort Reform in an Information Vacuum*, 37 EMORY L.J. 401, 402-03 (1988).

204. Malan v. Lewis, 693 P.2d 661, 669 (Utah 1984).

VI. CONCLUSION

If the Utah Supreme Court reviewed the Utah Health Care Malpractice Act's elimination of the disability tolling provision, the court would likely find it unconstitutional. Under the Utah Constitution's Open Courts Provision, Utah's minors have a well-established right to personally bring an action for negligent medical injury. The Malpractice Act abrogates that right but fails to provide an adequate alternative remedy. It also strikes an unreasonable balance between the restriction's questionable benefits and the complete abrogation of minors' longstanding protective disability status. Similarly, under Utah's Equal Protection Provision, the Malpractice Act's treatment of minors fails to substantially or reasonably minimize the adverse effect of rising malpractice insurance costs. Overall, the Malpractice Act's treatment of Utah's unrepresented children amounts to an unconstitutional infringement on their rights.

This conclusion is not intended to minimize the significant problems that exist with the current system of compensating malpractice victims and the negative impact it has on access to health care. Rather, the Utah Legislature will hopefully respond to this conclusion by reassessing the nature of the health care problem in the state and by examining the burden the Malpractice Act places on minors.

To the extent that malpractice costs ultimately restrict access to health care, the Utah Legislature should aggressively seek ways to reduce costs associated with unmeritorious claims and the substantial administrative costs of the compensation system.[205] To deal with the associated costs of malpractice claims by minors, for example, the Legislature might consider alternative compensation systems that eliminate expensive disputes over fault in cases involving severe birth injuries.[206] Additionally, the Legislature

---

205. *Cf.* AMA TASK FORCE, *supra* note 45, at 13 (goal of medical profession has been to change tort law in two general ways: "first, to make it harder for plaintiffs to bring groundless lawsuits, and second, to limit the impact of the costs of successful suits on defendants, insurers, and the medical profession as a whole").

206. For example, Virginia and Florida have developed no-fault programs to provide coverage for neurologically-impaired infants. *See* VA. CODE ANN. § 38.2-5000 (Michie Supp. 1990); FLA. STAT. ANN. § 766.301 (West Supp. 1992). The commentary on these programs has been mixed and their success unclear. *See, e.g.*, James A. Henderson, *The Virginia Birth-Related Injury Compensation Act: Limited No-Fault Statutes as Solutions*

should aggressively strive to improve access to health care. For example, the Legislature should try to secure obstetrical care in rural areas. State government can also foster and support private efforts to improve children's access to preventative medicine.[207]

On the other hand, the Utah Legislature should be wary of restricting the remedies available to minors. Clearly, society needs to prevent unrestrained growth in the costs associated with medical malpractice litigation. Additionally, tolling the statute of limitation for minors until the age of majority imposes unique costs on the malpractice insurance system. The Utah Constitution demands, however, that restrictions on such rights be carefully balanced with society's needs and that the means used can be proved to reasonably further those needs. Moreover, history counsels that when malpractice involves minor victims there are unique considerations that argue forcefully for unique treatment of minors under the law.

Accordingly, the Utah Legislature should consider alternative approaches to restricting a minor's right to personally bring a malpractice claim that better balance the conflicting interests. At a minimum, the Legislature should consider a tolling period for minors that spans their early years so they can reach an age when malpractice injuries are usually detectable, when injuries can be adequately evaluated, and when they can adequately communicate their injuries.[208] Additionally, the Legislature should distinguish between minors whose interests would likely be adequately

---

to the "Medical Malpractice Crisis," in 2 INSTITUTE OF MEDICINE, *supra* note 169, at 194 (questioning value of limited no-fault approach); Richard A. Epstein, *Market and Regulatory Approaches to Medical Malpractice: The Virginia Obstetrical No-Fault Statute*, in 2 COMMITTEE TO STUDY MEDICAL PROFESSIONAL LIABILITY AND THE DELIVERY OF OBSTETRICAL CARE, INSTITUTE OF MEDICINE, MEDICAL PROFESSIONAL LIABILITY AND THE DELIVERY OF OBSTETRICAL CARE 35-52, at 115, 131-32 (same). These attempts at least remain a starting point for evaluating and improving attempts to address the problems with the current compensation system. *See also* Randall R. Bovbjerg et al., *Obstetrics and Malpractice: Evidence on the Performance of a Selective No-Fault System*, 265 JAMA 2836 (1991) (applying a selective no-fault system based on "accelerated-compensation events" to 285 claims involving obstetrical and gynecological care and concluding system would settle claims faster and with less expense).

207. Walt Schaffer, *New Plan Helps Parents Insure Children*, SALT LAKE TRIB., Jan. 16, 1992, at B1.

208. *Cf.* TEX. REV. CIV. STAT. ANN. art. 4590i (West Supp. 1992) (minors under the age of twelve years shall have until their fourteenth birthday to file a claim).

represented by a parent or guardian and those whose interests would not.[209]

Restricting the rights of minors to seek a remedy for malpractice in an effort to address the problem of health care access for Utah citizens is a glaring example of an unconstitutional "arbitrary and impermissible shifting of collective burdens to individual citizens."[210] This is particularly true where the citizens burdened are minors who have lost a unique and well-deserved protective statute provided under Utah law for over 100 years.

ROB M. ALSTON

209. *Cf.* ARK. CODE ANN. § 16-114-203 (Michie Supp. 1991) (providing additional tolling of limitations period over age nine for minors without parent or guardian until they have a parent or guardian).

210. *See* Condemarin v. University Hosp., 775 P.2d 348, 358 (plurality opinion).

Guilty of Survival: *State v. Strieby* and Battered Women Who Kill in Utah

I. INTRODUCTION

In the last twelve years, battered women on trial for killing their abusive husbands or lovers have begun to plead self-defense. These cases manifest a commendable shift in black letter self-defense law toward accommodating individual variables that are relevant to determining whether a given homicide is justified. At the same time, these cases manifest unfortunate gender bias. The cases themselves are marked by stridently polarized majority and dissenting opinions, as is the scholarly commentary and press coverage.

State v. Strieby[1] is representative of the deeply divisive nature of many battered women's self-defense cases. In *Strieby*, the Utah Court of Appeals reversed Erlene Kay Strieby's manslaughter conviction arising out of the shooting death of her husband, Chris Strieby. The court held that the State had not met its burden of proving the absence of self-defense.[2] In a dissenting opinion, Judge Bench contended that the trial court had sufficient grounds for finding Strieby "could not have *reasonably* believed that shooting Mr. Strieby was necessary to prevent her own *imminent* death or serious bodily injury."[3] Both opinions speak the language of black letter law, but each belies deeper biases and concerns.

This Note begins in Section II by presenting a historical perspective on several of the gender biases that presently operate against battered women who kill. It then provides a quick overview of significant developments in the law of self-defense, as well as some polarized characterizations of battered women's self-defense cases. Section III examines Utah's response to battered women who kill and explores the implications of *State v. Strieby*.

---

1. 790 P.2d 98 (Utah Ct. App. 1990).

2. *Id.* at 101. The State must prove the absence of self-defense after it is properly raised. *See infra* note 186 and accompanying text.

3. *Strieby*, 790 P.2d at 102 (Bench, J., concurring and dissenting).

Finally, Section IV suggests guidelines for diminishing the impact of gender bias on battered women's self-defense claims.

II. GENDER BIAS AND BATTERED WOMEN'S SELF-DEFENSE CLAIMS

Battered women are hampered in their self-defense claims by deep-rooted societal biases. Two distinct biases operate against battered women who kill: traditional prohibitions against women taking autonomous, violent measures to protect themselves, especially against their husbands; and misconceptions regarding women who are victims in battering relationships. The following sections delineate these biases.

A. Violent Women in Historical Perspective

1. Traditional Views

Traditional notions of "the gentle sex" did not comport with women using violent measures to defend themselves. While men were esteemed for defending their honor or their person, women who did so were considered unfeminine.[4] Immanent in the female nature was "'all that is lovely in tenderness, affection and fidelity.'"[5] Violence was a man's province.[6]

---

4. *See* Bradwell v. State, 83 U.S. 130, 141 (16 Wall. 1872) (Bradley, J., concurring) ("Man is, or should be, woman's protector and defender."); ANGELA BROWNE, WHEN BATTERED WOMEN KILL 174 (1987) ("Traditional American attitudes encourage men to be active in defending their persons, families, and property; however, a woman is expected to have a husband, father, or son do it for her, or turn to the police for protection and help. . . ."); Elizabeth M. Schneider, *Equal Rights to Trial for Women: Sex Bias in the Law of Self-Defense*, 15 HARV. C.R.-C.L. L. REV. 623, 628 (1980) ("Women are discouraged from learning how to defend themselves because such behavior is 'unfeminine'; they are also taught to avoid engaging in violence.").

5. ANN JONES, WOMEN WHO KILL 87 (1980) (quoting from nineteenth-century trial of Lucretia Chapman).

6. *See id.* at 5-6. Jones discusses an 1897 study which argued that "'female criminals approximate more to males . . . than to normal women, especially in the superciliary arches in the seam of the sutures, in the lower jaw bones, and in the peculiarities of the occipital region.'" *Id.* at 6 (quoting CAESAR LOMBROSO & WILLIAM FERRERO, THE FEMALE OFFENDER 28 (New York, D. Appleton 1897)).

Wives who "rose up" in violence against their husbands were particularly offensive to traditional notions of womankind.[7] English common law designated the offense of a wife killing her husband as "a species of treason," a "much more atrocious crime" than that of a husband killing his wife.[8] Adhering to these traditional views, courts did not recognize that a woman's use of violence might be reasonable and justified. Women by nature were lacking in reason, ruled by emotion, susceptible to extreme nervous agitation, and only one step removed from hysteria.[9] Women were thought incapable of meeting the "reasonable man" standard the law required to justify killing in self-defense.[10] Thus, women who were acquitted of murder were "not thought to

---

7. *Id.* at 86-87. Jones details the nineteenth-century case of Ann Simpson, who was brought to trial for poisoning her husband. She won an acquittal by stressing the unimaginable nature of such a crime, for which there could be no motive:

> "A wife is arraigned to answer to the accusation of the foul and unnatural murder of her husband; charged with the cold destitution of all the finer feelings of our nature—a woman's nature: with the hellish and fiendlike spirit, which would prompt her, under the guise of love, to extend the poisoned chalice to him, to whom all her affections were pledged, all of her obedience due, and upon whom all her hopes of human happiness ought to have been centered. A crime so monstrous, so revolting, so unnatural, that one is tempted to pronounce its impossibility."

Id. at 100-02 (quoting from trial of Ann K. Simpson).

8. 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 418 n.103 (William D. Lewis ed., Philadelphia, Rees Welsh & Co. 1898) (1765). In his version of Blackstone's *Commentaries*, Professor Lewis included the following observation:

> Husband and wife, in the language of the law, are styled *baron* and *feme*. . . . [W]e might be inclined to think this merely an unmeaning technical phrase, if we did not recollect that if the baron kills his feme it is the same as if he had killed a stranger, or any other person; but if the feme kills her baron, it is regarded by the laws as a much more atrocious crime; as she not only breaks through the restraints of humanity and conjugal affection, but throws off all subjection to the authority of her husband. And therefore the law denominates her crime a species of treason, and condemns her to the same punishment as if she had killed the king. And for every species of treason . . . the sentence of women was to be drawn and burnt alive.

Id.

A colonial justice of the peace explained the special gravity of the crime of treason, stating: "This is a Crime of such transcendent Malignity, in its Nature and Consequences, that it is, by the Law, distinguished for its Superior Guilt, and therefore punishable in the most base and ignominious Manner: Other offenses are injurious to Private Persons only, but this is a Public Mischief, and often strikes at the Root of all Civil government." GEORGE WEBB, THE OFFICE AND AUTHORITY OF A JUSTICE OF PEACE 342-43 (photo. reprint 1969) (Williamsburg, William Parks 1736).

9. JONES, *supra* note 5, at 159-60.

10. Ronald K.L. Collins, *Language, History and the Legal Process: A Profile of the "Reasonable Man,"* 8 RUT.-CAM. L.J. 311, 315-20 (1977); Schneider, *supra* note 4, at 636.

have been exonerated," but rather to have "gotten away with something."[11]

2. *Modern Residua*

Traditional gender biases continue to operate against women who kill in self-defense. Modern literature persists in the notion that women who resort to violence are in conflict with their feminine nature.[12] Women are seen as more deranged than men who commit similar acts,[13] and are sentenced more harshly.[14] Moreover, society continues to expect women to retreat from confrontational situations rather than "manfully" stand their ground.[15] This expectation is reflected in judicial decisions,[16] and

---

11. JONES, *supra* note 5, at 8.

12. *See* RITA J. SIMON, WOMEN AND CRIME 2-9 (1975); Carol Smart, *Criminological Theory: Its Ideology and Implications Concerning Women*, *in* WOMEN AND CRIME IN AMERICA 6, 9-13 (Lee H. Bowker ed. 1981). One author notes that

> [a] woman who wields a deadly weapon, even to defend herself, presents a deeply disturbing image. Our prevailing ideal of women's true nature is that they are nurturers and life-givers, not life-takers. There is something that strikes us as obscene, against nature, or unholy about a woman who kills, that goes way beyond the illegality or even the immorality of her act. It shakes some of our most deeply held cultural stereotypes to their roots.

CYNTHIA K. GILLESPIE, JUSTIFIABLE HOMICIDE: BATTERED WOMEN, SELF-DEFENSE, AND THE LAW 12 (1989).

13. BROWNE, *supra* note 4, at 176; Elizabeth M. Schneider & Susan B. Jordan, *Representation of Women Who Defend Themselves in Response to Physical or Sexual Assault*, *in* WOMEN'S SELF DEFENSE CASES 1, 6 (Elizabeth Bochnak ed., 1981).

14. ELIZABETH BOCHNAK, WOMEN'S SELF-DEFENSE CASES 7; JONES, *supra* note 5, at 9; Schneider & Jordan, *supra* note 13, at 7; *see also* GILLESPIE, *supra* note 12, at 19 ("[W]omen who defend themselves against [violence] are charged with more serious crimes than men would be."). *But see* RITA J. SIMON & JEAN LANDIS, THE CRIMES WOMEN COMMIT, THE PUNISHMENTS THEY RECEIVE 57 (1991) ("Most observers feel that women receive preferential treatment, which in operational terms means that they are less likely than men to be convicted for the same type of offense; if they are convicted, they are less likely to be sentenced; and if they are sentenced, they are likely to receive milder sentences.").

15. BROWNE, *supra* note 4, at 174. Browne states:

> Although a majority of American jurisdictions now hold that there is no duty to retreat before using deadly force if the individual being attacked has reason to believe that the assailant will kill her or do serious bodily harm, society continues to expect that a woman *should* retreat, and the response is one of unease and censure when a woman takes action in her own defense.

Id.; *see also*, GILLESPIE, *supra* note 12, at 81-87 (criticizing modern presumption that wives should flee from own homes rather than defend themselves).

16. *See* People v. Lenkevich, 229 N.W.2d 298, 298 (Mich. 1975) (overturning second-degree murder conviction of woman defending herself against husband's in-home attack);

operates especially against women who stand their ground against their husbands.[17] Thus, even in jurisdictions that do not limit the defense of justifiable homicide by a duty to retreat, when a battered woman kills her abuser she may be required to show why she did not leave the relationship.[18]

In addition, the traditional notion that women are inherently unreasonable continues to operate against women who kill in self-defense. A person is justified in the use of force to defend against an attack if the person reasonably believes the force is necessary to protect against imminent danger of grave bodily harm.[19] Women have a difficult time demonstrating that their perceptions of danger are reasonable.[20] A study of twenty-seven women convicted of murder or manslaughter revealed that "[d]espite circumstances that could well ground a defense of self-defense, in only one case was self-defense argued."[21] Thus women have defended homicide charges by arguing mental impairment or diminished capacity. Until recently, these defenses were thought to stand a greater chance of success than the defense of justifiable homicide, based as they are on presumptions of women's mental weakness.[22]

People v. McGrandy, 156 N.W.2d 48, 49-50 (Mich. Ct. App. 1967) (overturning woman's manslaughter conviction for defending herself against husband's attack instead of taking back-door escape); State v. Grierson, 69 A.2d 851, 854-56 (N.H. 1949) (requiring woman to flee from home rather than defend against boyfriend's attack); *see also infra* text accompanying notes 227-29 (discussing trial court and dissenting opinions in *State v. Strieby*.

17. *See* People v. Jones, 12 Cal. Rptr. 777, 780 (Dist. Ct. App. 1961) ("The existence of the matrimonial status should be an additional reason to forego resort to a homicide.").

18. BROWNE, *supra* note 4, at 174 ("[T]he burden of proof usually falls on the woman to show why she could not leave the relationship, even though legally she need only demonstrate that she was in danger and was legitimately standing her ground in her own home."); Lenore E. Walker et al., *Beyond the Juror's Ken: Battered Women*, 7 VT. L. REV. 1, 5 (1982) ("Despite the legal right in most states to stand one's own ground on one's own property, women appear to be expected to have good reason for *not* leaving their homes or withdrawing from their marriages if they are under the threat of or actually experiencing assault.").

19. WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., CRIMINAL LAW 454-63 (2d ed. 1986); ROLLIN M. PERKINS & RONALD N. BOYCE, CRIMINAL LAW 1113-16 (3d ed. 1982).

20. *See* Schneider, *supra* note 4, at 636.

21. *Id.* at 638 n.82 (citation omitted). For a discussion of the ideological distinction between defenses grounded in excuse and defenses grounded in justification, see *infra* note 63.

22. *Id.* at 638; Schneider & Jordan, *supra* note 13, at 5. For a recent commentary advocating impaired mental state defenses for battered women who kill their abusers, see Rocco C. Cipparone, Jr., Comment, *The Defense of Battered Women Who Kill*, 135 U.

Indeed, even now, despite the shift in the past decade toward the use of self-defense in battered women's homicide cases, the perception of women as unreasonable continues to operate. Elizabeth Schneider argues convincingly that this perception is perpetuated in the form of battered woman syndrome evidence. The evidence, she contends, focuses on "passive, victimized aspects of battered women's experience, 'their learned helplessness,' rather than circumstances which might explain the homicide as a woman's necessary choice to save her own life."[23]

One of the most devastating consequences of the continued perception that women who kill cannot be justified in their behavior is the correlative perception that acquittal or lenient sentencing is not exoneration, but chivalry or mercy—an unequal application of the law in favor of women.[24] This perception is apparent in the criticism surrounding gubernatorial pardons of battered women sentenced to prison for killing their abusive spouses.[25] When Massachusetts Governor William Weld, on September 26, 1991, eased regulations under which battered women who murdered their abusers could petition for release from prison,[26] critics questioned Weld's singling out this group, "how-

---

PA. L. REV. 427, 429 (1987). *See also* Cathryn J. Rosen, *The Excuse of Self-Defense: Correcting a Historical Accident on Behalf of Battered Women Who Kill*, 36 AM. U. L. REV. 11, 18-24 (1986) (arguing battered women's self-defense claims should be treated as excuses rather than justifications). Conversely, Schneider & Jordan recommend that, rather than being relied on "almost automatically," an impaired mental state defense for a woman who commits homicide be thought of as a last resort, "with full awareness of its social implications." Schneider & Jordan, *supra* note 13, at 29.

23. Elizabeth M. Schneider, *Describing and Changing: Women's Self-Defense Work and the Problem of Expert Testimony on Battering*, 9 WOMEN'S RTS. L. REP. 195, 198-99 (1986).

24. For a discussion of the development and manifestation of the "chivalry" theory in diverse historical contexts, see JONES, *supra* note 5, at 8.

25. Over the past three years, the governors of Florida, Illinois, Louisiana, Maryland, New Hampshire, New Jersey, Ohio, Tennessee, and Washington have pardoned nearly three dozen battered women who claimed to have killed their abusers in self-defense. *See* Maria Puente, *Texas Considers Clemency; Will Review Cases Related to Abuse*, USA TODAY, May 17, 1991, at A3. In contemplation of granting clemency, the governor of Texas has directed special investigations of battered women's murder cases. *Id.* The governor of Massachusetts has "issued commutation guidelines which allow battered women seeking" clemency to argue that the abuse they suffered "significantly contributed" to their offenses. Frank Phillips, *Weld Relaxes Prison Appeal by Battered Women*, BOSTON GLOBE, Sept. 27, 1991, at 17. Advocacy groups in Arizona, California, Colorado, Kansas, and New York are pushing for clemency investigations similar to those in Texas. Puente, *supra*, at A3.

26. *See* Phillips, *supra* note 25, at 17.

ever deserving," for leniency. Comparing the governor's "throw-the-book" position toward drug dealers[27] with his "leniency" toward battered women, the critics intimated that battered women's homicides could not be justified. Similarly belying a reluctance to recognize battered women's homicides as justifiable self-defense, appeals for leniency often focus on *excusing* a woman who kills after incurring years of abuse, because she was victimized, "emotionally entangled," or had "suffered enough."[28]

B. Battered Women in Historical Perspective

1. Traditional Views

Battered women have suffered from societal norms and biases as malignant as those operating against women who kill. English common law granted husbands the right to "chastise" their wives, a privilege grounded in the rationale that "as [the husband] is to answer for her misbehavior, the law thought it reasonable to intrust him with this power of restraining her, by domestic chastisement."[29] For some "misdemeanors," a husband was privileged to "beat his wife severely with scourges and cudgels."[30] For others, only "moderate chastisement" was permissible.[31]

Early American cases upheld the husband's privilege of "domestic chastisement." In *Bradley v. State*,[32] a Mississippi court commented on a husband's conviction for assault and battery against his wife, stating: "[L]et the husband be permitted to exercise the right of moderate chastisement, in cases of great emergency, and use salutary restraints in every case of misbehav-

---

27. *Id.*

28. *See, e.g.*, Colman McCarthy, *Countering Violence at Home*, WASH. POST, July 23, 1991, at D13 (Ohio Governor Richard Celeste, first to grant clemency to a group of battered women, explained that "[t]hese women were . . . so emotionally entangled they were incapable of walking away."); *Battered Wife Freed in Killing of Husband*, N.Y. TIMES, Oct. 29, 1989, § 1, at 25 (Washington Governor Booth Gardner, giving his first clemency order in five years, released from prison a battered woman who had suffered "17 years of terror and abuse," stating that she had "suffered enough and society would not be threatened by her release.").

29. 1 WILLIAM BLACKSTONE, COMMENTARIES *444.

30. 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 417 n.101 (William D. Lewis ed. & trans., Philadelphia, Rees Welsh & Co. 1898) (1765).

31. *Id.*

32. 2 Miss. (1 Walker) 73 (1824).

ior, without being subjected to vexatious prosecutions . . . ."[33] Later courts concurred. The stated policy rationales included "preserv[ing] the sanctity of the domestic circle,"[34] closing the courthouse doors to "trivial complaints,"[35] and inducing the parties to "make the matter up and live together as man and wife should."[36] Not until 1871 did an American court deny a husband the right to beat his wife.[37] In *Fulgham v. State*,[38] the Alabama Supreme Court held that

> a rod which may be drawn through the wedding ring is not now deemed necessary to teach the wife her duty and subjection to the husband. The husband is therefore not justified or allowed by law to use such a weapon, or any other, for her moderate correction. The wife is not to be considered as the husband's slave. And the privilege, ancient though it be, to beat her with a stick, to pull her hair, choke her, spit in her face or kick her about the floor, or to inflict upon her like indignities, is not now acknowledged by our law.[39]

2. *Modern Residua*

Since Alabama's landmark ruling, the legal system's response to wife abuse continues to progress. The 1970s saw a wave of recognition accorded domestic violence.[40] Studies materialized

---

33. *Id.* at 74.

34. State v. Oliver, 70 N.C. 60, 61 (1874). The court preferred to "draw the curtain, shut out the public gaze, and leave the parties to forget and forgive." *Id.* at 61-62.

35. *Id.* at 61.

36. State v. Black, 60 N.C. (Win.) 262, 263 (1864).

37. *See* INFORMATION AIDS, INC., DOMESTIC VIOLENCE—NO LONGER BEHIND THE CURTAINS 52 (Mark A. Siegel et al. eds., 1987).

38. 46 Ala. 143 (1871).

39. *Id.* at 146-47.

40. Victoria M. Mather, *The Skeleton in the Closet: The Battered Woman Syndrome, Self-Defense, and Expert Testimony*, 39 MERCER L. REV. 545, 548 (1988); Kathleen Waits, *The Criminal Justice System's Response to Battering: Understanding the Problem, Forging the Solutions*, 60 WASH. L. REV. 267, 268 n.2 (1985); Terry L. Fromson, Note, *The Case for Legal Remedies for Abused Women*, 6 N.Y.U. REV. L. & SOC. CHANGE 135, 136 (1977); Sarah B. Vandenbraak, Note, *Limits on the Use of Defensive Force to Prevent Intramarital Assaults*, 10 RUT.-CAM. L.J. 643, 643 (1979).

The movement toward recognizing the devastating consequences wreaked by wife abuse achieved momentum in the United States in 1975, when the National Organization for Women established a Task Force on Battered Women and Household Violence. *See* Virgie L. Mouton, Note, *Wife Abuse Legislation in California, Pennsylvania, and Texas*, 7 T. MARSHALL L. REV. 282, 282 n.1 (1982).

that revealed the devastating toll spouse abuse takes on society.[41] Domestic violence proved to be not only widespread, but also dangerous, with some two to four thousand women being beaten to death annually.[42] Women in battering relationships were shown to be in perhaps the highest risk group for homicide, with forty-one percent of all women killed being killed by their spouses.[43]

Once the devastating nature of wife abuse was revealed, judges, legislators, and law enforcement officers began to acknowledge its criminality.[44] Efforts were made, and continue to be made, to provide women more adequate protection against their abusers.[45]

Notwithstanding heightened awareness of domestic violence, and concomitant reform efforts, traditional misconceptions remain. Many judges and law enforcement officers (as well as potential jurors and society at large) continue to believe that men have the right to "chastise" their wives for their "misbehavior,"[46] or that battering incidents are merely "quarrels" between husband and wife that should be worked out privately.[47] These beliefs rest on

---

41. *See* INFORMATION AIDS, INC., *supra* note 37.

42. Jane O'Reilly, *Wife Beating: The Silent Crime*, TIME, Sept. 5, 1983, at 23 (cover story).

43. Mather, *supra* note 40, at 546 ("FBI statistics showed that husbands or boyfriends killed 40% of all female homicide victims."); Schneider, *supra* note 4, at 626; Marvin E. Wolfgang, *A Sociological Analysis of Criminal Homicide*, *in* STUDIES IN HOMICIDE 15, 23 (Marvin E. Wolfgang ed., 1967).

The proportion of women killed by men is greater than that of men killed by women. A 1956 study of 588 homicide victims found that 41% of the women victims were killed by their husbands as compared to only 11% of the male victims who were killed by their wives. *See* Marvin E. Wolfgang, *Husband-Wife Homicides*, 2 J. SOC. THERAPY 263-271 (1956). Moreover, of the husbands killed by their wives, a large proportion were "victim-precipitated." *Id.* at 268-69. According to an estimate from Cook County, Illinois, 40% of all women held there on homicide charges were accused of killing a man who had battered them. Schneider & Jordan, *supra* note 13, at 7.

44. Waits, *supra* note 40, at 268-69.

45. Mather, *supra* note 40, at 559-60.

46. LENORE E. WALKER, THE BATTERED WOMAN 12 (1979). In many states the decision whether to arrest and prosecute a wife batterer is based on the number of stitches her wounds need, indicating a continued tolerance for husbands' "moderate chastisement" of their wives. *Id.*

47. *See id.* at 13. Studies have demonstrated a widespread unwillingness to intervene in violence between spouses. One study found that witnesses intervened when a woman was being assaulted by another woman, or when a man was being assaulted by a woman; however, witnesses did not intervene when a woman was being assaulted by a man, inferring that the woman was being assaulted by her husband. Murray A. Straus, *Sexual Inequality, Cultural Norms, and Wife-Beating*, 1 VICTIMOLOGY 54, 61

a variety of presumptions. One is that if the abuse were not trivial, the woman would leave the relationship. Another is that if the woman stays, she must like, need, or deserve the abuse.[48] Still another is that wives provoke their husbands into beating them.[49]

Perhaps manifesting these persisting beliefs and presumptions, the criminal justice system, despite recent progress, still fails to treat violence perpetrated against battered women as a crime.[50] A recent study found that only ten percent of police departments serving populations of 100,000 or more encourage arrest in domestic assault cases.[51] In addition, prosecutors rarely pursue charges against batterers.[52] By one estimate, only one domestic violence case in one hundred ever reaches a courtroom.[53]

(1976).

Batterers, reflecting the belief that what a husband does to his wife is a private matter, often express outrage at any interference from the criminal justice system in the way they run their homes. *See* WALKER, *supra* note 46, at 12.

48. Schneider, *supra* note 4, at 625. There is a widely held belief that battered women are masochistic. More than 40% of the respondents in one study felt that a woman was masochistic if she stayed in a relationship after having been beaten. *See* Charles P. Ewing & Moss Aubrey, *Battered Woman and Public Opinion: Some Realities About the Myths*, 2 J. FAM. VIOLENCE 257, 261 (1987).

Perhaps the first proponent of the theory that women tend toward masochism was Sigmund Freud. *See* Waits, *supra* note 40, at 280 n.56. More recent proponents of the theory have concluded that beatings give women "masochistic gratification," helping them "deal with the guilt arising from the intense hostility expressed in [their] controlling, castrating behavior." John E. Snell et al., *The Wifebeater's Wife*, 11 ARCHIVES GEN. PSYCHIATRY 107, 111 (1964).

49. James Ptacek, *Why Do Men Batter Their Wives?*, *in* FEMINIST PERSPECTIVES ON WIFE ABUSE 133, 154 (Kersti Yellö & Michelle Bograd eds., 1988). One study found that 20% of randomly selected respondents in a telephone interview survey reported that women usually cause the beatings they receive. Karen M. Gentemann, *Wife Beating: Attitudes of a Non-clinical Population*, 9 VICTIMOLOGY 109, 109 (1984). Another study asked 216 people a variety of questions relating to a hypothetical battering incident. In response to a question asking whether the woman was in part responsible for her husband's assaultive behavior, 38.3% said yes. Ewing & Aubrey, *supra* note 48, at 261.

50. Ptacek, *supra* note 49, at 155.

51. *Arrest May Be Deterrent in Domestic Violence, Study Shows*, N.Y. TIMES, May 30, 1984, at C4. Most domestic assaults "are treated by the police and the courts as misdemeanors, even where quite serious injuries have been inflicted." GILLESPIE, *supra* note 12, at 136. In misdemeanor cases, police officers can only make an arrest if the act is committed in their presence. Thus, arrest rates are typically quite low in cases of interspousal assault. DEL MARTIN, BATTERED WIVES 90 (1976).

52. *See* Elizabeth Stanko, *Fear of Crime and the Myth of the Safe Home*, *in* FEMINIST PERSPECTIVES ON WIFE ABUSE 75, 79 (1988).

53. GILLESPIE, *supra* note 12, at 142. In Detroit in 1972, warrants were prepared in

Moreover, judges continue to trivialize the danger a battered woman faces. One judge chastised a woman for obtaining a restraining order against her husband, who had beaten her, choked her, and threatened to kill her, indicating that she was taking up the court's time when "it has a lot more serious matters to contend with."[54] She had requested a police escort to accompany her to the apartment she shared with her husband so she could pack some clothes. The judge refused, saying: "[Y]ou don't need the police . . . [just] go there and act as an adult."[55] Less than one month later, the woman's body was found, having been shot, stabbed, and strangled to death by her husband.[56]

Other incidents demonstrate the continuing presumption that women may be beaten or killed for "provoking" their husbands with "misbehavior." One judge sentenced a man, who abused his wife for fifteen years and ultimately shot her five times in the face, to two years in a work-release program. The judge stated that leniency was appropriate because the man's wife had engaged in certain "highly provoking acts" before her death.[57] These acts included deceiving her husband by acting very loving toward him immediately before fleeing the marriage, and failing to leave him a note explaining her departure.[58] In another such display of traditional bias, a New York City police sergeant, after hearing the details of a brutal assault, noted that "[m]aybe she wasn't giving him what he needed sexually."[59] Thus, presumptions regarding the nature of domestic violence lead to toleration of violence perpetrated against women by the men with whom they are intimately involved.

The criminal justice system's unresponsiveness to battered women's complaints may have lethal consequences. Several commentators argue that failure to intervene effectively in

---

4900 domestic assault cases. Of these cases, fewer than 300 were tried in court. MARTIN, *supra* note 51, at 114.

54. Eileen McNamara, *Judge Criticized After Woman's Death,* BOSTON GLOBE, Sept. 21, 1986, Metro Section, at 1.

55. *Id.*

56. *Id.*

57. Jennifer Parmalee, *'We're Outraged,'* ST. PETERSBURG TIMES, June 25, 1983, at 1A.

58. *Id.*

59. O'Reilly, *supra* note 42, at 23.

domestic violence contributes to interspousal homicide.[60] One study revealed that more than eighty percent of women killed by their spouses sought police help one to five times before being killed.[61] Thus, as one commentator notes, "until improved policies are adopted and take effect, a battered woman's first contact with the legal system may be as a criminal defendant."[62]

C. The Law of Self-Defense

Self-defense is the most widely recognized defense to intentional homicide. It rests on the belief that a person is justified in defending him or herself against physical harm.[63] Generally, individuals may resort to deadly force to protect themselves if they reasonably believe that they are in imminent danger of death or serious bodily harm, and deadly force is necessary to avoid the harm.[64] This definition gives rise to four issues commonly the focus of jurisprudence surrounding battered women's self-defense claims: (1) the standard for measuring the reasonableness of the defendant's actions; (2) the temporal proximity of danger facing the defendant; (3) the proportionality of force used to meet the threatened harm; and (4) the defendant's duty to retreat.[65]

---

60. *See, e.g.*, BOCHNAK, *supra* note 14, at 10-11 ("[I]nadequate protection has serious consequences for women, since . . . incidents of domestic violence commonly result in serious physical injury or death for the woman."); Michael Dowd, *Dispelling the Myths About the "Battered Woman's Defense:" Towards a New Understanding* 567, 571 (1992) ("[B]attered women have known that they could expect little protection from a society made up of individuals who resembled, at least in thought, the men who beat them. Some of these women have died as a result, but others, in the face of impending death, have fought back and killed their abusers."); M.D.A. Freeman, *Domestic Violence: The Limits of Effective Legal Action*, 20 CAMBRIAN L. REV. 17, 22 (1989) (summarizing research linking pro-arrest policies with downward trend in domestic homicides); Margaret Howard, *Husband-Wife Homicide: An Essay from a Family Law Perspective*, 49 L. & CONTEMP. PROBS. 63, 87 (1986) (encouraging changes in legal system in order that pattern of escalation that may lead to homicide will be interrupted).

61. WALKER, *supra* note 46, at 64.

62. Cipparone, *supra* note 22, at 428.

63. Schneider, *supra* note 4, at 630. Successfully invoking self-defense pardons an individual's use of force because the circumstances justify it. Acting in self-defense is thus held to be correct, or even "laudable" conduct. Phyllis L. Crocker, *The Meaning of Equality for Battered Women Who Kill Men in Self-Defense*, 8 HARV. WOMEN'S L.J. 121, 130-31 (1985); Schneider, *supra* note 4, at 630-31. Self-defense is ideologically distinct from defenses that excuse the wrongful use of force because the individual's state of mind was such that the conduct is tolerated. *See* Crocker, *supra*, at 130-31.

64. LAFAVE & SCOTT, *supra* note 19, at 454.

65. *See* Holly Maguigan, *Battered Women and Self-Defense: Myths and Misconcep-*

1. Reasonable Belief

The reasonableness of a defendant's belief in the necessity of using deadly force in self-protection is measured by varying standards. Some jurisdictions employ an objective standard of reasonableness and a few employ a subjective standard. The majority employ a standard that incorporates both objective and subjective elements.[66]

Strict adherence to an objective standard of reasonableness may impede battered women's self-defense claims.[67] Social mores, based on deeply rooted societal values and priorities, determine when conduct is "objectively" reasonable.[68] As we have seen, the traditional status of a married woman vis-a-vis her husband was such that a wife killing her husband would not have been considered reasonable under any circumstances.[69]

While women's status in society appears to be improving,[70] widely held beliefs about the impropriety of a woman defending herself violently against her lover or spouse remain.[71] Specifically, beliefs that continue to hamper a battered woman's ability to establish the reasonableness element of her self-defense claim include: (1) the belief that women are inherently irrational;[72] (2) the belief that a man who has beaten his wife before without killing her would be unlikely to kill her in subsequent beatings;[73]

---

tions in Current Reform Proposals, 140 U. PENN. L. REV. 379, 385 (1991).

66. *Id.* at 457.

67. Walter W. Steele, Jr. & Christine W. Sigman, *Reexamining the Doctrine of Self Defense to Accommodate Battered Women*, 18 AM. J. CRIM. L. 169, 176 (1991).

68. Schneider, *supra* note 4, at 635.

69. *See supra* note 8 and accompanying text (discussing common law's treatment of women who killed their husbands); *see also* Crocker, *supra* note 63, at 131 n.48 ("To see women's acts of self-defense as justified—particularly when they result in the death of a husband—is to turn Blackstone on his head. The argument posits the act as legally justifiable where once it was treasonous. . . . This may illuminate the fundamental difficulty with the legal system's response to women's self-defense claims.").

70. *See* Norman J. Finkel et al., *The Self-Defense Defense and Community Sentiment*, 15 L. & HUM. BEHAV. 585, 593 (asserting that community sentiment, as gauged by mock juror studies, shows considerable support for not guilty by reason of self-defense verdicts in battered woman cases). *But see* GILLESPIE, *supra* note 12, at 94 ("Where self-defense is involved, . . . it is difficult to avoid the conclusion, looking at convictions, that jurors frequently are unwilling to believe that it is *ever* reasonable for a woman to kill her mate.").

71. GILLESPIE, *supra* note 12, at 129-30.

72. Schneider, *supra* note 4, at 636.

73. *See* GILLESPIE, *supra* note 12, at 24, 59-60. In People v. Chapman, 364 N.E.2d

and (3) the belief that battered women enjoy or provoke the beatings they receive.[74] Jurisdictions that fail to incorporate subjective elements into their reasonableness standard run the risk that beliefs such as these will inform jurors' "objective" account of reasonableness.

Fortunately, at present the majority of jurisdictions employ a mixed subjective-objective standard.[75] These jurisdictions allow the biases operating against battered women's self-defense claims to be countered by expert testimony on the battered woman syndrome.[76] The testimony, based largely on research initiated by Dr. Lenore Walker,[77] attempts to explain the reasons battered women remain in their relationships other than supposed masochistic desire or inexplicable female irrationality. One expert explains that the "testimony begins with a description of what studies have shown to be characteristic of battered women. . . . An important aspect of this portion of the testimony comes from the fact that research findings often go against prevailing misconceptions about battered women. Thus, misconceptions can be corrected."[78] Moreover, the testimony elucidates the cyclical nature of the violence battered women experience[79] and its escalation over time.[80] It thus demonstrates that as the beatings increase in severity, a woman's belief that this time she will be killed may well be reasonable.[81]

---

577 (Ill. App. Ct. 1977), the court relied on this reasoning to uphold a battered woman's manslaughter conviction. *Id.* at 580-81. The defendant introduced evidence of prior beatings. *Id.* at 580. "The evidence," the court found, "does not establish any reason which would require the court to believe that this assault was more serious than the others." *Id.*

74. GILLESPIE, *supra* note 12, at 157.

75. *See* Maguigan, *supra* note 65, at 409; Steele & Sigman, *supra* note 67, at 176. Only a distinct minority of jurisdictions currently employ either the pure objective or subjective standards. *See* LAFAVE & SCOTT, *supra* note 19, at 457-58; Maguigan, *supra* note 65, at 409.

76. *See* Steele & Sigman, *supra* note 67, at 176. A healthy majority of jurisdictions now allow expert testimony in battered women's self-defense cases. *See* Lenore E. Walker, *Battered Women Syndrome and Self-Defense*, 6 NOTRE DAME J. L. ETHICS & PUB. POL. 321, 321 (1992); Jeanne-Marie Bates, Comment, *Expert Testimony on the Battered Woman Syndrome in Maryland*, 50 MD. L. REV. 920, 920-21 (1991).

77. WALKER, *supra* note 46; *see also* State v. Williams, 787 S.W.2d 308, 311 (Mo. Ct. App. 1990) (citing Walker as originator of battered woman syndrome theory).

78. JULIE BLACKMUN, INTIMATE VIOLENCE: A STUDY OF INJUSTICE 190 (1989).

79. *See* WALKER, *supra* note 46, at 55-70.

80. GILLESPIE, *supra* note 12, at 59-60.

81. Cipparone, *supra* note 22, at 434.

In addition, whether an objective, subjective, or mixed standard of reasonableness is employed, a battered woman may introduce evidence of her batterer's past acts of violence against her. This evidence, sometimes referred to as "history of violence" evidence, is admissible in every jurisdiction to establish the reasonableness of a defendant's fear of danger.[82] A battered woman may also introduce evidence that her batterer threatened her with death or violence to explain the reasonableness of her, what otherwise might be perceived as, overly severe self-protective measures. In *People v. Bush*,[83] for example, a California Court of Appeals held that a person whose life has been threatened is justified in acting more quickly and taking harsher protective measures than would be a person not receiving such threats.[84]

Thus, in many jurisdictions self-defense law as it now stands does not unfairly prohibit a battered woman from establishing that her perception of danger was reasonable. Evidence of her batterer's prior violence toward her, and perhaps his threats of death, coupled with expert testimony aimed at dispelling anachronistic notions about battered women who kill, will in many cases be sufficient to meet the combined subjective-objective standard.

2. *Imminent Danger*

The law of self-defense, besides requiring that the defendant's apprehension of danger be reasonable, requires that a threatened attack be imminent before force is used to defend against it.[85] Underlying this requirement is the rationale that, where there is no certainty that a threat of future violence will in fact be carried

---

82. *See* JOHN W. STRONG, MCCORMICK ON EVIDENCE § 193 (4th ed. 1992); Maguigan, *supra* note 65, at 421 n.145, 422; Susan Estrich, *Defending Women*, 88 MICH. L. REV. 1430, 1436 (1990) (book review). The decedent's history of violence toward third persons, if known to the defendant, is admissible under the same theory of relevancy. *See* LAFAVE & SCOTT, *supra* note 19, at 457; Maguigan, *supra*.

83. 148 Cal. Rptr. 430 (Ct. App. 1978).

84. *Id.* at 435-37; *see also* State v. Spaulding, 257 S.E.2d 391 (N.C. 979) (holding actual show of deadly force unnecessary where assailant who made threats and advanced with hand in pocket had stabbed defendant on prior occasion); State v. Wanrow, 559 P.2d 548 (Wash. 1977) (reversing defendant's conviction because jury instructed not to consider defendant's knowledge of assailant's reputation for violence, but only acts or circumstances "at or immediately before the killing").

85. LAFAVE & SCOTT, *supra* note 19, at 458.

out, homicide cannot be condoned.[86] Traditionally, a threat made without a *present* ability to carry it out would not satisfy the imminence requirement, no matter how certain it was that the threatened conduct would eventually occur.[87] Even a threat made with the present ability to carry it out had to be coupled with an overt act in order to be reasonably perceived as imminent.

Jurisdictions adhering to narrow imminence standards limit the focus of inquiry to the particular instant of defendant's action. Jurisdictions applying a broader standard, on the other hand, allow inquiry into the circumstances, including past events, surrounding the defendant's resort to self-defense.[88] Further, some jurisdictions require prima facie evidence of an overt act before expert witness testimony or history of violence evidence can be introduced. Others allow the defendant to introduce expert witness testimony or history of violence evidence in order to establish the presence of an overt act.[89]

Many battered women would have difficulty meeting the imminence requirement under the narrow or prima facie overt act standards. Battered women kill within the context of an ongoing and escalating cycle of violence.[90] They may come to perceive that danger is imminent based on their intimate familiarity with their batterer's cycles of behavior—from loving and conciliatory, to tense, to violent. As the battered woman lives through increasingly violent episodes, her perception that her death is imminent becomes increasingly acute. She must be allowed to show that, based on her experience with her batterer, she reasonably perceived a seemingly insignificant gesture to be an "overt act" signalling the imminent onslaught of deadly violence.[91] If this evidence is not presented, a battered woman's claim that, for instance, she knew her batterer was about to kill her because of the tone of his voice, would seem incredible.

---

86. *Id.*; Cipparone, *supra* note 22, at 437.

87. GILLESPIE, *supra* note 12, at 67; Vandenbraak, *supra* note 40, at 651-52; *see also* People v. Lucas, 324 P.2d 933, 936 (Cal. Dist. Ct. App. 1958) ("[T]hreats alone, unaccompanied by some act which induces in defendant a reasonable belief that bodily injury is about to be inflicted, do not justify a homicide.").

88. Maguigan, *supra* note 65, at 414.

89. *Id.* at 424.

90. WALKER, *supra* note 46, at 61-62.

91. *Id.*

Most jurisdictions recognize that inflexible adherence to the narrow or prima facie overt act requirement is often inappropriate.[92] These jurisdictions allow expert witness testimony to explain how a battered woman, having experienced numerous tension-building phases that erupt into violence at the slightest instigation, and having thus been conditioned into hypervigilance to her batterer's every move, might reasonably perceive impending danger or an overt act where a jury of twelve objective persons might not.[93] Courts allowing this evidence have held that a very subtle or insignificant act on the part of the batterer,[94] or a change in the batterer's pattern of abuse,[95] may be sufficient to constitute a reasonable perception on the part of a battered woman that she is in imminent danger of death or serious bodily harm. Thus, under the law of many jurisdictions, the imminence requirement does not present the insurmountable difficulty to battered women's self-defense claims that it might once have.

3. Proportional Force

The third often-litigated element of self-defense law is that of proportional force. The proportional force requirement limits the use of deadly force to stave off an attack to those instances in which deadly force is itself threatened.[96] If rigidly applied, this rule would preclude the use of a deadly weapon against an

---

92. Maguigan, *supra* note 65, at 424. For a breakdown of jurisdictions applying narrow or broad temporal proximity standards, as well as traditional versus lower-showing overt act requirements, see Professor Maguigan's table, *id.* at 461.

93. Walker, *supra* note 76, at 324; *see also* GILLESPIE, *supra* note 12, at 135 ("The threats and verbal harangues and 'little batterings' of the tension-building stage that lead so inexorably to a major beating may seem relatively trivial in themselves. But when they are recognized as part of a pattern they can be seen for what they are . . . : signals that worse violence is to come. Since the woman knows from her own painful experience that the next beating may well be worse than the last and that once it starts she will be powerless to stop it, . . . [w]hat might appear at first glance to be an overreaction is an entirely reasonable response to the situation she actually faces . . . .").

94. *See* State v. Hodges, 716 P.2d 563, 565, 569 (Kan. 1986); State v. Osbey, 710 P.2d 676, 678-80 (Kan. 1985); State v. Hundley, 693 P.2d 475, 476, 479-80 (Kan. 1985).

95. State v. Williams, 787 S.W.2d 308, 312-13 (Mo. Ct. App. 1990).

96. LAFAVE & SCOTT, *supra* note 19, at 456. The self-defense provisions of some states extend lawful use of deadly force to extreme intrusions on freedom of the person, such as kidnapping and rape. *Id.* at 456 n.15. Utah is among these states. *See infra* note 155 (setting forth Utah's justifiable homicide provision, UTAH CODE ANN. § 76-2-402(1) (Supp. 1992), which provides that deadly force is justified "to prevent the commission of a forcible felony").

unarmed assailant.[97] Since the vast majority of battered women use weapons—usually guns—against their batterers, who most often use only their hands,[98] the rule would preclude self-defense claims in most cases involving battered women.

However, an overwhelming majority of jurisdictions do not adhere to a rigid proportional force requirement.[99] Rather, account is taken of the respective size and sex of the assailant and defendant, and of the anticipated nature of the unarmed attack in light of a history of abuse or other factors relevant to the defendant's assessment of risk.[100] Thus a battered woman, under the majority approach to the proportional force rule, should not be precluded from a self-defense defense because she used a gun against her unarmed attacker.

4. Duty to Retreat

The majority of jurisdictions do not require persons under attack to retreat, even if they can do so safely, rather than stand their ground against an aggressor.[101] In the "strong" minority that do impose a duty to retreat, "taking what might be regarded as a more civilized view,"[102] a person is not required to retreat from their home except, in some jurisdictions, when the assailant is a co-occupant. This co-occupant exception to the minority rule may work to the disadvantage of some women's self-defense claims since it most battering incidents occur in premises battered women share with their batterers.[103] However, this aspect of the duty to retreat is rarely a determinative issue in battered women's self-defense cases.[104]

The more formidable aspect of the duty to retreat for battered women has been its confusion with a supposed obligation to leave

97. LAFAVE & SCOTT, *supra* note 19, at 456.
98. Maguigan, *supra* note 65, at 416 n.131.
99. *Id.* at 417.
100. LAFAVE & SCOTT, *supra* note 19, at 456-57.
101. *Id.* at 460.
102. *Id.*
103. Having reviewed 270 battered women's homicide cases, Professor Maguigan places the proportion at between 56% and 61% depending on whether the battering incident took place in the context of a confrontational or a non-confrontational case. Maguigan, *supra* note 65, at 420 n.141.
104. *Id.* at 420.

the relationship. Commentators[105] and trial courts[106] make this mistake, though few appellate courts do.[107] When it occurs, it can be explained only as a misunderstanding of the law.[108]

D. Polar Approaches

Based on this summary overview of modern self-defense principles, it appears that many instances in which battered women kill their abusers would meet the substantive requirements of self-defense. Yet applications of the law of self-defense that operate to vindicate battered women have met with some resistance, and battered women's self-defense claims in general are a subject of considerable controversy. On one side of the controversy are contentions that the law of self-defense is being applied more leniently to battered women.[109] On the other are contentions that the law of self-defense is inherently biased against women and that equal protection mandates that those biases be addressed. The remainder of this section outlines the basic arguments advanced in support of both positions.

1. Women are Getting Lenient Treatment

Gubernatorial pardons, lenient sentences, and acquittals of women who murder abusive spouses have resulted in the perception that women are "getting away with murder."[110] A battered

---

105. *See, e.g.*, Stephen J. Schulhofer, *The Gender Question in Criminal Law*, 7 SOC. PHIL. & POL'Y 105, 129 (1990) ("Many women do successfully escape from abusive mates. Those who instead resort to deadly force should have to prove the concrete circumstances that prevented them from doing likewise."); *see also* sources cited *infra* note 134.

106. *See, e.g.*, cases cited *supra* note 16.

107. Maguigan, *supra* note 65, at 419.

108. *See infra* text accompanying notes 217-19 (discussing dissenting judge's imputation of duty to retreat on Erlene Strieby in derogation of Utah law).

109. *See, e.g.*, Mira Mihajlovich, Comment, *Does Plight Make Right: The Battered Woman Syndrome, Expert Testimony and the Law of Self-Defense*, 62 IND. L.J. 1253, 1269, 1281 (1986-87) (asserting that testimony of battered woman syndrome neutralizes the law of self-defense); Marilyn H. Mitchell, Note, *Does Wife Abuse Justify Homicide*, 24 WAYNE L. REV. 1705, 1729-30 (1978) (discussing "unwise expansion of homicide defenses"); Estrich, *supra* note 82 (asserting battered woman syndrome evidence should not be admitted to eliminate imminence requirement).

110. GILLESPIE, *supra* note 12, at 3-10; *see, e.g.*, Geraldine Baum, *Should These Women Have Gone Free?*, L.A. TIMES, Apr. 15, 1991, at E1 (outlining criticism of Maryland governor's grants of clemency to eight battered women who killed or assaulted

woman's history of abuse is perceived as giving her a license to kill, to engage in vigilante justice, or to declare an "open season on men."[111] The legal arguments advanced against battered women's self-defense claims reflect these perceptions.

(a) Empathy Legally Insufficient to Justify Murder

Battered women's self-defense pleas have been called "a request to abandon limits on self-defense out of empathy for the circumstances of the defender and disgust for acts of [the] abuser."[112] This characterization of battered women's homicide cases views the introduction of evidence regarding the history of prior abuse as solicitation for sympathy.[113] Proponents of this view insist that sympathy must not operate to "carve[] out an exception to general principles and sanction[] homicide for a particular class of persons."[114]

---

their mates); Tamar Lewin, *Criticism of Clemency May Affect Efforts to Free Battered Women*, N.Y. TIMES, Apr. 2, 1991, at A17 (setting forth political obstacles to governors granting clemency); Cathy Young, *Getting Away with Murder*, NEWSDAY, June 18, 1991, at 42 ("[M]any people . . . find it unsettling that a woman, even one who has been horribly abused, should be able to kill a man in his sleep or hire a hit man almost with impunity.").

111. JONES, *supra* note 5, at 289-92. The first trial to receive national attention was that of Francine Hughes, who was charged with first degree murder in the March 9, 1977, death of her husband. One of Mrs. Hughes's neighbors told reporters at the time of her trial that "[i]f she gets out of this there'll be a lot of dead guys lying around." *Id.* at 290. Another commented that her acquittal would mean "open season on men." *Id.; see also* Kay Bartlett, *Spousal Homicide Law: "Open Season" on Men—or Domestic Violence?*, L.A. TIMES, Mar. 17, 1991, at A33 (prosecutors protested Ohio Governor Richard Celeste's grant of clemency to 26 battered women, saying he might just as well have declared "open season on men"); Richard Cohen, *Vigilante Justice Back in Women's Movement; Women Turn the Tide on Vigilante Justice*, WASH. POST, Dec. 4, 1977, at B1 (feminist response to Hughes case is adoption of worst of male-dominated code—"the notion that for certain kinds of behaviour a rough, vigilante justice is in order"); Susan Diesenhouse, *Women Driven to Kill are Shown More Mercy*, N.Y. TIMES, Jan. 30, 1989, at A10 (opponents of gubernatorial pardons noted that being battered is not an excuse for "preplanned executions"); Jerald K. Footlick & Elaine Sciolino, *Wives Who Batter Back*, NEWSWEEK, Jan. 30, 1978, at 54 (battered women's murder cases reveal a trend that "smacks uncomfortably of frontier justice" and might create "a new legitimacy for violent retaliation").

112. Estrich, *supra* note 82, at 1437.

113. *See* Mitchell, *supra* note 109, at 1726 ("Although unstated, the real plea in abused wife homicide cases seems to be 'I couldn't take anymore.' That has never been a legal defense.").

114. *Id.* at 1731.

The Supreme Court of North Carolina adopted this view in *State v. Norman*.[115] Judy Norman, the defendant, killed her husband of twenty-five years, J.T., while he napped.[116] J.T. had forced Judy to prostitute herself, beat her when she refused and when she did not bring in enough money, and had threatened to kill or maim her on several occasions.[117] The court noted that the defendant's evidence was "poignant."[118] However, the evidence did not justify "stretching the law of self-defense to fit the facts of this case."[119] Such a result "could not be limited to a few cases decided on evidence as poignant as this."[120]

The Supreme Court of Kansas reached a similar conclusion in *State v. Stewart*.[121] In that case, Peggy Stewart killed her husband Mike while he slept.[122] Mike had threatened to kill Peggy on several occasions. Once, he shot her pet cat and then put the gun to Peggy's head and threatened to pull the trigger. He ordered Peggy to kill her daughter and bury her. When Peggy ran away, Mike retrieved her, told her he would kill her if she ever ran away again, and forced her to have oral sex with him several times. The next morning, he indicated that she need not bother cleaning house because she would not be there long, and that she should not bother with her things because she could not take them with her. Peggy shot Mike later that evening as he slept.

The court sympathized with the notion that Mike Stewart was an "evil man who deserved the justice he received from his battered wife."[123] However, the court held that to allow a self-defense instruction would modify "the law of self-defense to be more generous to one suffering from the battered woman syndrome than to any other defendant relying on self-defense."[124]

---

115. 378 S.E.2d 8, 15 (N.C. 1989).
116. *Id.* at 9-11.
117. *Id.*
118. *Id.* at 15.
119. *Id.*
120. *Id.*
121. 763 P.2d 572, 579 (Kan. 1988).
122. The facts are taken from the court's opinion in *Stewart*, 763 P.2d at 574-76.
123. *Id.* at 579.
124. *Id.*

(b) Vigilante Justice

Battered women's use of deadly force against their abusers has been characterized as "a kind of vigilante justice the law is supposed to preclude."[125] This view sees battered women's use of deadly force as a reaction to threats of future abuse, or retaliation for past abuse. If the law were to justify such behavior, the argument goes, it would be sanctioning "instant execution" of all wife beaters.

The majority adopted this view in *Stewart*.[126] The opinion characterized Peggy's killing of Mike as revengeful, noting Peggy's reply when asked why she had killed her husband: "It was as if Mike was going to do something again like had been done before."[127] Thus, the court concluded that Peggy was not entitled to a self-defense instruction.[128] "To hold otherwise," the court stated, "would in effect allow the execution of the abuser for past or future acts and conduct."[129] Further, "'[t]o permit capital punishment to be imposed upon the subjective conclusion of the [abused] individual that prior acts and conduct of the deceased justified the killing would amount to a leap into the abyss of anarchy.'"[130]

The *Norman* court similarly characterized Judy Norman's killing of her husband as motivated by revenge or retaliation.[131] The court brought out evidence that Judy had expressed considerable anger toward her husband and said she would kill him "because of the things he had done to her."[132] The court held that the requirements for self-defense could not be "relaxed" to accommodate these facts. To do so, the court stated,

> would tend to categorically legalize the opportune killing of abusive husbands by their wives solely on the basis of the wives' testimony concerning their subjective speculation as to the probability of future felonious assaults by their husbands.

---

125. Mitchell, *supra* note 109, at 1731.
126. *Stewart*, 763 P.2d at 579.
127. *Id.* at 575.
128. *Id.* at 579.
129. *Id.*
130. *Id.* (quoting Jahnke v. State, 682 P.2d 991, 997 (Wyo. 1984)).
131. *See* State v. Norman, 378 S.E.2d 8, 9-11 (N.C. 1989).
132. *Id.* at 10-11.

> Homicidal self-help would then become a lawful solution, and perhaps the easiest and most effective solution, to this problem.[133]

(c) She Was Free to Leave

The threshold argument for opponents of battered women's self-defense claims is that "a woman in a violent relationship would appear always to have an alternative to killing the man—leaving the relationship."[134] The majority opinions in both *Norman* and *Stewart* emphasized this aspect of the evidence. In *Norman*, the court stated that "the defendant had ample time and opportunity to resort to other means of preventing further abuse by her husband."[135] Similarly, the *Stewart* court noted that Peggy "voluntarily agreed to return home with Mike," and that she "showed no inclination to leave."[136] Further, "Peggy's annulment and divorces from former husbands, and her filing for divorce after leaving Mike, proved that Peggy knew there were non-lethal methods by which she could extricate herself from the abusive relationship."[137]

2. Self-Defense Law Is Biased Against Women

(a) Equal Treatment, Not Empathy

The central distinction between proponents and opponents of battered women's self-defense claims is the perception of the law's treatment of battered women: opponents perceive that the law affords battered women special treatment,[138] while proponents perceive that the law treats battered women less favorably than

---

133. *Id.* at 15.

134. David L. Faigman, Note, *The Battered Woman Syndrome and Self-Defense: A Legal and Empirical Dissent*, 72 VA. L. REV. 619, 621-22 (1986); *see also* Note, *The Effect of Marriage on the Rules of the Criminal Law*, 61 COLUM. L. REV. 73, 94 (1961) ("Abusive conduct may be sufficiently controlled by the spouse's . . . ultimate remedy of leaving the home.").

135. *Norman*, 378 S.E.2d at 13.

136. *Stewart*, 763 P.2d at 578.

137. *Id.* at 576.

138. *See, e.g.*, Mitchell, *supra* note 109, at 1731 ("[A]bused wives who retaliate with deadly force against an abusive husband appear to be getting favored treatment from judges and juries in derogation of the law.").

it does men.[139] The proponent's argument is two-pronged: (1) the elements of self-defense presuppose circumstances far removed from those actually present when battered women kill their batterers,[140] and (2) traditional biases operate to make the law of self-defense discriminatory in effect.[141] Ignoring these aspects of a battered woman's self-defense claim, proponents contend, "places a burden on her that her male counterpart is not asked to bear."[142]

In a landmark decision, styled *State v. Wanrow*,[143] the Washington Supreme Court recognized that to apply blindly the elements of self-defense to the circumstances in which a woman defendant kills would deny her equal treatment under the law of self-defense. The court found that the trial court's instructions on the law of self-defense could have left the jury with the impression that the standard of reasonableness was one "applicable to an altercation between two men."[144] Such a standard would ignore the fact that the defendant was a five-foot-four-inch woman who, as a result of sex discrimination, had no training or experience in physical combat.[145] Her perceptions of danger could not be expected to be the same as those of a "reasonable man" involved in a physical struggle with an equally matched assailant. Thus, she could reasonably perceive the need to use a weapon to defend herself, even though her assailant was not armed.[146] The court warned:

> Until such time as the effects of [our nation's long and unfortunate history of sex discrimination] are eradicated, care must be taken to assure that our self-defense instructions afford women the right to have their conduct judged in light

---

139. *See, e.g.*, Schneider, *supra* note 4, at 630-38 (arguing sexual prejudices restrict women's self-defense claims).

140. *See* GILLESPIE, *supra* note 12, at 5.

141. *See* GILLESPIE, *supra* note 12, at xiii, 191; *see also* Crocker, *supra* note 63, at 132 (suggesting individualized approach "can equalize the positions of male and female defendants by recognizing their differences"); Schneider, *supra* note 4, at 640 (stating greater individualization necessary to provide equal treatment for battered women raising self-defense claims).

142. Schneider, *supra* note 4, at 640.

143. 559 P.2d 548 (Wash. 1977).

144. *Id.* at 558.

145. *Id.*

146. *See id.*

> of the individual physical handicaps which are the product of sex discrimination. To fail to do so is to deny the right of the individual woman involved to trial by the same rules which are applicable to male defendants.[147]

In addition to contending that the presuppositions of self-defense law may not apply when women defend themselves against men, proponents of battered women's self-defense claims argue that traditional biases affect a battered woman's ability to successfully present her case.[148] To receive equal treatment, a battered woman must be able to present evidence explicitly addressing the inapplicability of stereotypical attitudes in the circumstances of her case.[149] Many courts have accepted that expert testimony may be introduced for such edification purposes.[150]

(b) Reasonable Response

Opponents of battered women's self-defense claims characterize the women's actions as retaliatory, or responses to mere *threats* of future harm, while proponents attempt to show that battered women act in response to real danger, which they perceive based on long experience with their abuser's patterns of abuse. Proponents argue that a battered woman's experience with her abuser gives her insight into the risk she faces that may well make her action reasonable.[151]

In *State v. Gallegos*,[152] the New Mexico Court of Appeals accepted this characterization of reasonableness. The lower court had denied Anita Gallegos a self-defense instruction because it

---

147. *Id.* at 559. Courts have responded favorably to *Wanrow*. The modern trend in the law of self-defense is to consider individual circumstances that might bear on the reasonableness of a defendant's defensive actions, such as the size of the defendant vis-a-vis her attacker, and the attacker's reputation for violence or prior history of violence with the defendant. *See* Mather, *supra* note 40, at 565.

148. Crocker, *supra* note 63, at 132; *see supra* note 71 and accompanying text.

149. *See Crocker*, *supra* note 63, at 132.

150. *See* Mather, *supra* note 40, at 574.

151. *See* Estrich, *supra* note 82, at 1436. Estrich argues that a woman's choice may be reasonable if the woman is probably right in her assessment. "Does the woman know something we don't about the risk she faces? If the answer is yes, then the jury should know it as well, and take it into account." *Id.*

152. 719 P.2d 1268 (N.M. Ct. App. 1986).

found that there had been no "obvious threat at the time of the slaying."[153] However, Anita had been threatened by her husband George on several prior occasions. George had threatened to cut off her breasts if they grew any larger and had placed a gun to her head and said he would kill her if she ever left him. On several occasions, he tied her hands behind her back and sodomized her until she bled. On the day Anita killed George, he had once again sodomized her and threatened to kill her if she left him. When he ordered her into the bedroom, she responded by taking a loaded rifle in with her and shooting him at close range while he lay in bed. The court emphasized that, based on "her knowledge of what had happened to her in similar circumstances," Anita's perception of an imminent threat of danger arising when George called her into the bedroom may well have been reasonable.[154] Thus, she was entitled to a self-defense instruction. "To deny the defense of self-defense under the facts of this case," the court concluded, "would ignore reality."[155]

(c) No Alternatives

Finally, proponents of battered women's self-defense claims characterize the battering relationship as one in which the woman is trapped, rather than one in which she is free to leave. This argument gains much of its force from empirical evidence demonstrating the legal system's failure to adequately respond to battered women's pleas for help.[156] Moreover, proponents contend, the abuse itself is often such that the battered woman becomes a hostage in her own home. When she leaves, he searches for her, finds her, and brings her back.[157] He threatens to kill her if she leaves again.

---

153. *Id.* at 1269. The facts are taken from the court's opinion in *Gallegos*, 719 P.2d at 1271-72.

154. *Id.* at 1272.

155. *Id.* at 1273.

156. *See* Mather, *supra* note 40, at 556-59. One study demonstrating the inefficacy of police protection in Kansas City found that 90% of the city's interfamily homicides were preceded by at least one call to the police, while five or more calls were made in 50% of the cases. *See id.* at 557.

157. Women who leave their abusive spouses are in extreme danger. Batterers search for their wives or girlfriends and escalate the abuse, or kill them, when they find them. *See* MARTIN, *supra* note 51, at 78-79.

This is the view the dissent emphasized in *State v. Norman*.[158] The opinion noted that when Judy Norman left her husband, he retrieved her; when she went to the Department of Social Services, he retrieved her.[159] Witnesses testified as to Judy's experiences with social service agencies and the law, "which had contributed to her sense of futility and abandonment through the inefficacy of their protection and the strength of her husband's wrath when they failed."[160] The day before Judy killed her husband she called the police. The officer responding to her call later testified that her face was bruised, that she was crying, and that she said her husband had been beating her all day until she could not take any more. The officer told her he could do nothing unless she took out a warrant on her husband. She responded that if she did, her husband would kill her.[161] This evidence, coupled with the defendant's testimony that she felt her husband was inescapable, was enough to convince the dissenting judge that Judy was not "free to leave." He concluded that "[w]here torture appears interminable and escape impossible, the belief that only the death of the oppressor can provide relief is reasonable in the mind of a person of ordinary firmness."[162]

These cases demonstrate the difficulty battered women's self-defense claims pose. Invoking self-defense appears inapposite when a woman shoots her husband while he lies resting, or in response to a request that she come into the bedroom. Yet the circumstances of these cases, seen from a different perspective, seem to compel a self-defense instruction. Perhaps much of the controversy would be quelled were both sides to recognize that substantive self-defense law supports the majority of battered women's self-defense claims. Thus there need be no struggle either for or against a major overhaul in carefully developed principles of justification. A more appropriate struggle might instead focus on whether available doctrines are being uniformly applied. But in the meantime, polar characterizations of fact-specific events leading up to the ultimate act of homicide are sure to continue.

---

158. 378 S.E.2d 8, 16-17 (N.C. 1989) (Martin, J., dissenting).
159. *Id.* at 17 (Martin, J., dissenting).
160. *Id.* at 18 (Martin, J., dissenting).
161. *Id.* at 19 (Martin, J., dissenting).
162. *Id.* at 18 (Martin, J., dissenting).

III. UTAH'S RESPONSE TO BATTERED WOMEN WHO KILL

A. Case Studies

At least eight Utah women have killed abusive husbands or live-in boyfriends since 1987.[163] The cases of two of these women, Nancy Riley, who pled guilty to negligent homicide,[164] and Patricia Hughes, released June 23, 1992, from the Utah State Prison after serving time for attempted manslaughter,[165] are discussed below. Erlene Strieby's case is discussed in section III.B below.

Nancy Riley's relationship with her boyfriend, James Gentry, began to turn sour six months after he moved into the house she shared with her children from a former marriage—that is when the beatings and the sexual abuse began.[166] James started by demanding sex with Nancy on the living room couch while her children watched television.[167] This progressed to demanding sex with her while her children watched. James then began to include Nancy's son in the abuse, putting a sheet over Nancy's head and forcing her son to perform sexual acts on her. If Nancy protested she was beaten. Her injuries from James' beatings had on several occasions required emergency medical care. Nancy summoned the police several times to intervene on her behalf, but the police were unable to prevent the abuse from continuing.

On June 5, 1987, James's sexual demands took on yet a new dimension. He told Nancy that he was going to initiate her 13-year-old daughter into his sexual intrigues, and demanded that Nancy bring her to him. Nancy left the room where James lay on the couch, returned with a 20-gauge shotgun, and shot James point-blank in the back of the head.

---

163. *Abused Women Who Killed Their Husbands*, SALT LAKE TRIB., July 21, 1991, at A1.

164. Nancy Hobbs, *Abuse Victim: My Terror Excited Him*, SALT LAKE TRIB., July 21, 1991, at A1.

165. Telephone Interview with June Hinkley, Records Officer, Utah State Prison (Oct. 14, 1992) [hereinafter Hinkley Interview]; Nancy Hobbs, *Beatings Ended When Soft-Spoken Wife Grabbed Knife*, SALT LAKE TRIB., July 21, 1991, at A4.

166. *See* Hobbs, *supra* note 165.

167. The facts discussed here are taken from an interview with Brooke C. Wells, Defense Attorney for Nancy Riley (Nov. 13, 1991) [hereinafter Wells Interview].

Nancy was charged with manslaughter.[168] The criminal complaint, filed June 19, 1987,[169] alleged that Nancy shot James "under the influence of extreme emotional disturbance, for which there was a reasonable explanation or excuse."[170] On the advice of her attorney, Nancy told the police and prosecutors working on her case of the circumstances surrounding James's murder.[171] Her story was amply corroborated by records from several different hospitals detailing her physical abuse, by police records of domestic violence calls made from her house, by the testimonies of her son and daughter, and by the testimonies of other witnesses who had either witnessed James's abuse of Nancy or had seen the bruises and other injuries on Nancy's body after the beatings.[172]

Both the prosecution and defense agreed that Nancy should not be incarcerated.[173] Yet, her actions seemed to fall outside the realm of Utah's self-defense statutes.[174] Nancy created a "'prose-

---

168. *Magna Woman, 30, Says Innocent in June Killing of Boyfriend,* SALT LAKE TRIB., Aug. 22, 1987, at B2.

169. *Hearing Set for Magna Woman Charged In Boyfriend's Death*, SALT LAKE TRIB., July 10, 1987, at B3.

170. *Id.; see also* UTAH CODE ANN. § 76-2-402 (Supp. 1992) (defining "self-defense").

171. Wells Interview, *supra* note 167.

172. *Id.*

173. Some of the detectives involved in Nancy's case felt that, given the extenuating circumstances, charges should not have been filed against her. Mike Carter, *Abused Lover Ends Nightmare*, SALT LAKE TRIB., July 10, 1987, at B1. Nancy's defense attorney indicated that had Nancy not hidden James's body for two days after the murder, charges would not have been filed. Wells Interview, *supra* note 167. Bud Ellett, chief of the Salt Lake County Attorney's Justice Division, stated that had circumstances been even slightly different, charges may not have been brought. *See* Carter, *supra*.

174. Nancy was charged under Utah's manslaughter statute, which provides a legal excuse for killing "under the influence of extreme emotional disturbance for which there was a reasonable explanation or excuse"; or "under circumstances where the actor reasonably believes the circumstances provide a legal justification or excuse for his conduct although the conduct is not legally justifiable or excusable under the existing circumstances." UTAH CODE ANN. § 76-5-205(1)(b), (c) (1990).

> Utah's justifiable homicide (self-defense) statute provides:
> A person is justified in threatening or using force against another when and to the extent that he reasonably believes that such force is necessary to defend himself *or a third person* against such other's imminent use of unlawful force; however, a person is justified in using force which is intended or likely to cause death or serious bodily injury only if he reasonably believes that the force is necessary to prevent death or serious bodily injury to himself or a third person, *or to prevent the commission of a forcible felony*.

UTAH CODE ANN. § 76-2-402(1) (Supp. 1992) (emphasis added).
Subsection (3) provides that "a forcible felony includes . . . rape, forcible sodomy, and aggravated sexual assault." *Id.* § 76-2-402(3).

cutorial dilemma.'"[175] On the one hand, there was the feeling that "women who find themselves in this position [should] leave it to the proper authorities" rather than take the law into their own hands.[176] On the other hand, there was an uneasy awareness that Nancy had *tried* to "leave it to the proper authorities," and the authorities had been unable to help her.

In the end, the charge against Nancy was amended to a class A misdemeanor count of negligent homicide.[177] In a hearing before Third District Judge Richard Moffat, Nancy pled guilty and was placed on "good behavior" probation.[178] Deputy Salt Lake County Attorney (now Third District Judge) Leslie Lewis told Judge Moffat that the plea bargain was appropriate because Nancy had "suffered a great deal through this," and "[w]hile the state cannot condone this behavior, I believe I understand the reasoning that went behind the act."[179] Judge Moffat agreed with the amended charge, commenting that while Nancy had placed herself in the abusive situation, she had "paid a very dear price for [it]," and "that price paid is enough."[180] He told Nancy, "I am hopeful you can straighten your life out."[181]

Patricia Hughes was not shown as much sympathy. She received a zero-to-five-year sentence for attempted manslaughter in the stabbing death of her husband Charles.[182] Charles was in the habit of beating Patricia, but usually Patricia felt that she could stop him from "going too far."[183] On the day she killed him,

---

175. Carter, *supra* note 173 (quoting statement of Bud Ellett, chief of the Salt Lake County Attorney's Justice Division).

176. *Id.* James's murder was the second incident that month in which a woman had "taken the law into her own hands" and killed an abusive spouse. *Id.* Peggy Brown shot her husband Bradley while he slept, one month before Nancy shot James. Charges were not filed against Peggy. Sheriff's Homicide Sergeant Garth Beckstead indicated that the shooting occurred after Peggy and her children had been assaulted by Bradley. According to the sheriff's logs, deputies had been called to the Brown residence 8 to 12 times the prior year for domestic violence. *Id.* Prosecution against Peggy was not pursued, at least in part because of a question whether her *Miranda* rights had been violated in an interrogation that took place in the sheriff's office after the shooting. *See id.*

177. Mike Carter, *Woman Gets Probation For Killing Abusive Boyfriend,* SALT LAKE TRIB., April 9, 1988, at B1.

178. *Id.*

179. *Id.*

180. *Id.*

181. *Id.*

182. Hinkley Interview, *supra* note 165.

183. *See* Hobbs, *supra* note 165.

however, the beating was such that she did not know whether she would live or die.[184] It began when she returned from work. Charles accused her of hiding his whiskey, which she denied.[185] He then picked up a skillet and smashed her in the head.[186] When she collapsed to the ground he began kicking her where she lay, saying, "I'm going to kill you bitch, I'm going to kill you."[187] After "sustain[ing] enough punishment to finally fight back," Patricia struggled for a butcher knife and stabbed Charles.[188] Her daughter, who witnessed the incident, claims that either Patricia or Charles would have been dead before the fight ended.[189]

Patricia was initially charged with second-degree murder. The charge was reduced to attempted manslaughter when Patricia produced evidence of previous beatings Charles had inflicted on her. She pled guilty and did not argue self-defense.[190] Patricia was released from prison June 23, 1992.[191]

Both these cases indicate an unwillingness to recognize abused women's self-defense claims. Nancy was shown mercy; Patricia was not. Yet neither woman was seen as responding in a reasonable, justifiable manner to defend herself (or, in Nancy's case, her daughter) against the grave danger confronting her.

Utah's self-defense statute provides that a person is justified in using deadly force if "he reasonably believes the force is necessary to prevent death or serious bodily injury to himself or

---

184. *Id.*
185. *Id.*
186. *Id.*
187. *Id.*
188. *Id.*
189. *Id.*
190. Patricia's defense attorney, Nancy Bergeson, stated that Patricia was not a typical battered woman because she did not accept her husband's physical assaults passively. Telephone Interview with Nancy Bergeson, Federal Public Defender, Portland, Or. (November 23, 1991). One commentator has noted that court opinions have relied on deviations from the battered woman stereotype to deny self-defense instructions. *See* Crocker, *supra* note 63, at 144-50. In Patricia's case, the perception that a woman must be a "typical battered woman" in order to argue self-defense may have impeded her defense.

Patricia may also have felt reluctant to argue self-defense based on her abuse at Charles's hands because she had previously served time for attempted distribution of a controlled substance. Hinkley Interview, *supra* note 165. Her criminal record would have made her an unsympathetic victim.

191. Hobbs, *supra* note 165.

a third person, or to prevent the commission of a forcible felony."[192] The statute defines forcible felony to include "rape, forcible sodomy, and aggravated sexual assault."[193] This definition of justifiable homicide seems to cover Nancy's actions. In light of Nancy's past experiences with James, in which Nancy had been unable to prevent his sexual abuse of herself or her son, she could reasonably have believed that deadly force was necessary to prevent the rape of her daughter.[194]

The facts in Patricia Hughes' case appear to compel, to an even greater extent than those in Nancy Riley's, a finding of justifiable homicide. There can be no question that Patricia responded to imminent danger: Charles was beating her over the head with a skillet and kicking her while she lay on the floor. To find it unreasonable for a person to respond in such circumstances with deadly force seems to contradict the meaning of justifiable homicide.

B. State v. Strieby

1. Facts

In *State v. Strieby*,[195] the Utah Court of Appeals reversed Erlene Kay Strieby's manslaughter conviction for the shooting death of her husband Chris.[196] Chris and Erlene had a history of violent encounters with each other.[197] On July 8, Erlene went to Chris's place of employment to tell him she was leaving him. They argued, and Chris told Erlene she "'was a cunt and not to come around him anymore.'"[198] He gave her the finger, whereup-

192. UTAH CODE ANN. § 76-2-402(1) (Supp. 1992).

193. *Id.* § 76-2-402(3).

194. Nancy's defense attorney felt that Nancy had a good case for self-defense. However, Nancy did not wish to risk a conviction, since she had children depending on her, and so opted to accept the plea bargain. Wells Interview, *supra* note 167.

195. 790 P.2d 98 (Utah Ct. App. 1990).

196. *Id.* at 101. Erlene served 16 months in the Utah State Prison before her conviction was reversed. Nancy Hobbs, *I'm Guilty Only of Survival, Says Acquitted Killer*, SALT LAKE TRIB., July 21, 1991, at A4.

197. *See* Hobbs, *supra* note 196. The facts are taken from the court's opinion in *Strieby*, 790 P.2d at 99, from the account Erlene gave reporters for the Salt Lake Tribune in Hobbs, *supra* note 196, and from briefs filed for appellant and respondent in *Strieby*.

198. Brief of Appellant at 3, State v. Strieby, 790 P.2d 98 (Utah Ct. App. 1990) (No. 890124-CA).

on she left and went to a bar and had several drinks. Later in the afternoon, she went to pick Chris up, and again they argued. Chris pinned her to the floor and held her there by the neck for a few minutes while she struggled. When he released her, Erlene called a friend to take her home.

Later that evening Chris arrived home. Immediately upon entering the front door he began yelling obscenities at Erlene and threatening to kill her. He said, "'I'll beat you to death. No wonder your first husband beat you. You're a mouthy bitch.'"[199] Erlene told Chris to leave her alone and that if he would give her a few days to pack she would move out. He said, "'I ain't giving you no time at all.'"[200] She tried to escape out the front door, but Chris slammed the door shut, telling her that "'two other women had left him, and the only way [she] could leave was on a stretcher.'"[201] They struggled in the entryway, and then Erlene attempted to run up the stairs. Chris caught her by the foot, dragged her down the stairs, and began hitting her with his fists. She dodged his punches and was able to fend him off with a kick to his groin. Chris retreated to the kitchen and Erlene ran up the stairs to their bedroom.

Chris kept a .357 magnum under his bed. Erlene got the gun and returned to the top of the stairs to meet Chris as he ascended, shouting obscenities and threatening to kill her. Erlene testified that she "had never seen him like this. 'I had never seen his eyes—I had never seen his face contorted to the point where he was completely uncontrollable, where he—I couldn't talk to him.'"[202] Chris "'just—kept coming up the stairs. He just kept coming. And he just kept screaming at me that he could get me before I could shoot him.'"[203] Erlene pointed the gun at him and told him to let her pack and leave. Chris continued up the stairs, still threatening to kill her. Erlene fired a single shot, killing Chris instantly.

Erlene was charged with second-degree murder. She waived her right to a jury trial, choosing to argue her position before Third District Court Judge Kenneth Rigtrup. At trial, the state's

199. *Id.* at 5.
200. *Id.*
201. *Id.* at 9.
202. *Id.* at 10.
203. *Id.*

case-in-chief consisted of the testimony of a neighbor, police officers, and the state medical examiner.[204] The only evidence the state adduced as to what occurred during the violent encounter was a taped statement Erlene gave Detective Alan James and Sheriff Don Proctor following the shooting as she sat in the sheriff's car at the scene.[205] At the close of the state's case-in-chief, defense counsel made a motion for judgment of acquittal. Counsel argued that the state's case introduced evidence of self-defense, which, once raised, the state had the burden of disproving beyond a reasonable doubt. This, the defense argued, the state had not done. Judge Rigtrup denied the motion.[206]

After hearing the defense case, which consisted of Erlene's testimony expanding on her voluntary statement at the scene and the testimony of the physician who examined Erlene at the emergency room following the shooting, Judge Rigtrup found Erlene guilty of manslaughter. He acknowledged that there were no substantial conflicts in the evidence and that, given Chris's "powerful, muscular build," he had "no substantial doubt about the reality" of Erlene's fear for her life.[207] Nevertheless, Judge Rigtrup found that the violence was "perhaps not of the magnitude as perceived by Mrs. Streiby,"[208] and concluded that Erlene's actions did not constitute self-defense because: (1) Erlene was "quicker and more agile" than Chris, not being as drunk as he was, and thus Erlene could have eluded him;[209] and (2) Chris's retreat to the kitchen, where he fixed himself a drink,[210] indicated a "reasonable, substantial cessation" in the hostilities.[211] Erlene appealed her conviction to the Utah Court of Appeals.

---

204. *Id.* at 2-7.

205. *Id.* at 7. The taped statement recounted the events leading up to and following the shooting much the same as do the facts set forth *supra*, notes 197-203 and accompanying text, and in the appellate court's opinion. Erlene's quoted statements in the text accompanying notes 198-201 are taken from the statement taken immediately subsequent to the shooting. All other quoted material is taken from either Erlene's statements at trial or her statements to newspaper reporters.

206. Brief of Appellant at 7, *Strieby* (No. 890124-CA).

207. *Id.* at 17; *Strieby*, 790 P.2d at 101.

208. Brief of Respondent at 23, *Strieby* (No. 890124-CA).

209. *Id.*

210. The evidence conflicted as to whether Chris fixed himself a drink while in the kitchen. *Strieby*, 790 P.2d at 101.

211. *Id.*

2. Holding and Reasoning

The court of appeals overturned Erlene's manslaughter conviction, finding it "contrary to the clear weight of the evidence."[212] The court began by briefly reiterating Erlene's testimony that Chris's violent physical attack, escalated beyond anything she had previously experienced with him, coupled with his threats to kill her, led her to believe that force was necessary to protect herself from death or serious bodily injury.[213] The trial judge's finding that there was "no substantial doubt" that Erlene feared for her life obviated any need to inquire into whether *actual* belief existed. The only issue before the court, then, was whether Erlene *reasonably* believed that deadly force was required. In reaching its conclusion that her belief was in fact reasonable, the court addressed only the two issues raised by the trial court: (1) whether Chris was capable of seriously harming Erlene, given his level of intoxication, and (2) whether there had been a substantial cessation in the hostilities.[214]

With regard to the first issue, the court concluded that the evidence sufficiently illustrated that "Chris was fully capable of seriously harming defendant" at the time she killed him.[215] Specific evidence on which the court relied included medical testimony that Chris's level of intoxication was not high enough to strongly affect his coordination, Erlene's testimony that Chris had been able to prevent her from escaping out the front door, and medical testimony describing the serious injuries Chris inflicted on Erlene immediately before she killed him.[216]

The court next addressed whether there had been a cessation in the hostilities prior to the killing. Emphasizing that "[i]t is not the defendant's burden to 'establish a defense of self-defense beyond a reasonable doubt, or even by a preponderance of the

212. *Id.* The court also found that Judge Rigtrup had properly denied defense counsel's motion for judgment of acquittal at the close of the State's case-in-chief. The court held that the State's evidence, which included Erlene's admission that she shot Chris, was sufficient to establish the prima facie elements of manslaughter since the "'[a]bsence of self-defense is not one of the prima facie elements of homicide.'" *Id.* at 100 (quoting State v. Knoll, 717 P.2d 211, 214 (Utah 1985)).

213. *Id.* at 100-01.

214. *See id.*

215. *See id.*

216. *See id.*

evidence,"[217] the court held that the supposition of a cessation was merely a "'speculative leap[] across . . . remaining gap[s] in the evidence.'"[218] The court concluded that the prosecution had not sufficiently established that there had been a substantial cessation in the hostilities. It stated:

> All the evidence . . . indicates that defendant reasonably believed herself to be in imminent danger of serious injury or death and that she had, indeed, already suffered serious injury. The trial court did not doubt her credibility, but nevertheless engaged in pure speculation about a cessation of hostilities. While the trial court's conclusions should be respected, the conviction may not oppose the weight of the evidence.[219]

3. Dissent

Judge Bench dissented from the majority opinion, emphasizing that Utah law only justifies the use of deadly force if a person is subject to imminent attack, and the person reasonably believes deadly force is necessary to prevent death or serious bodily injury.[220] The circumstances, Judge Bench reasoned, supported the trial court's conclusion that Erlene "could not have *reasonably* believed that shooting Mr. Strieby was necessary to prevent her own *imminent* death or serious bodily injury."[221] Whether Chris fixed himself a drink or not, Judge Bench noted, the evidence was uncontroverted that Chris had gone into the kitchen after Erlene kicked him in the groin. This was sufficient to indicate a cessation of hostilities. At that point, Erlene "could have done any number of things to protect herself. Instead, she went up stairs, picked up a loaded revolver, and waited for Mr. Strieby."[222] In a footnote, Judge Bench quoted the trial court's statement that:

---

217. *Id.* (quoting State v. Knoll, 712 P.2d 211, 214 (Utah 1985)). Once the issue of self-defense is raised, the prosecution has the burden of proving beyond a reasonable doubt that the defendant did not act in self-defense. *See* UTAH CODE ANN. §§ 76-1-501, -502 (1990).

218. *See Strieby*, 790 P.2d at 101 (quoting State v. Petree, 659 P.2d 443, 445 (Utah 1983)).

219. *Id.*

220. *Id.* at 102 (Bench, J., dissenting).

221. *Id.* (Bench, J., dissenting).

222. *Id.* (Bench, J., dissenting).

> "[t]here was no physical evidence that Mrs. Strieby had, other than initially, tried any vigorous efforts to escape the condo, to shout, to holler, to seek the assistance of others. There was no evidence in the record to suggest that she had used objects of furniture or anything else in the house to fend him off or to take any evasive action. But rather that she went up the stairs and got the gun at a time when he wasn't in vigorous, hot pursuit."[223]

4. Analysis

The *Strieby* opinions manifest polarized characterizations of the facts and law that obtain generally in battered women's self-defense cases.[224] Judge Bench's dissenting opinion characterized Erlene as vindictive, stating that, instead of exercising any number of options open to her, Erlene "went upstairs, picked up a loaded revolver, and waited for Mr. Strieby."[225] He also quoted a witness's testimony that Erlene had said, prior to the shooting, that "as soon as she got her hands on a gun, he was a dead SOB."[226]

Further, both Judges Rigtrup and Bench argued that Erlene's conduct was unreasonable because "she could have done any number of things to protect herself" and there was no evidence that she tried to escape (other than her own testimony that Chris prevented her from escaping out the front door), or that she sought the assistance of others before taking defensive measures. Though couched in the language of reasonableness, the argument implies that Erlene had a duty to retreat from her home.

This is not the law in Utah. In *State v. Starks*,[227] the defendant, after an altercation in a restaurant, got a gun and returned to the scene of the fight.[228] Nevertheless, the Utah Supreme Court held that the defendant was entitled to a self-defense instruction. According to the court:

---

223. *Id.* at 102 n.1 (Bench, J., dissenting).

224. *See supra* Section II.D. (discussing polar characterizations of battered women's self-defense claims).

225. *Strieby*, 790 P.2d at 102 (Bench, J., dissenting).

226. *Id.* at 102 n.1 (Bench, J., dissenting).

227. 627 P.2d 88 (Utah 1981).

228. *Id.* at 89.

> That defendant in this case armed himself and went to a location where he knew he would find the deceased does not of itself deprive him of his right to self-defense. One is entitled to go where he has a right to be without losing his right to assert self-defense in a murder prosecution.[229]

Thus Erlene need not have attempted to escape from her home before resorting to self-defense. Nor need she have elicited the assistance of others to aid in her protection. The trial court and dissenting opinion's mischaracterization of the law perhaps belies their unspoken belief that women should not protect and defend themselves, but wait to be protected and defended.

Judge Rigtrup's cessation of hostilities theory further evinces an extraordinary unwillingness to see Erlene's violent self-protection as reasonable. Evidence was presented at trial that lying next to Chris's dead body when the police arrived was a blue plastic cup and a damp spot.[230] From this, the judge speculated that Chris must have fixed himself a drink after Erlene kicked him in the groin and he retreated to the kitchen. While Chris fixed himself a drink, the theory seems to be that Erlene should have made her escape. Judge Bench echoes this basis for the emphasis on the blue cup, stating, "[w]hether he fixed a drink or not, I believe the evidence supports the trial court's finding regarding cessation of hostilities. At that point [after Chris went into the kitchen], defendant could have done any number of things to protect herself."[231]

Judge Bench's point is well taken. But, absent a duty to retreat, the escape options available to Erlene are irrelevant to a finding that her resort to deadly force in self-defense was legally justified. Moreover, in the famous words of Justice Holmes, "[d]etached reflection cannot be demanded in the presence of an uplifted knife."[232] Erlene need not have done the best thing, or the most moral or humane thing. She need only have done one of the perhaps many reasonable things that the circumstances permitted.

---

229. *Id.* at 91.

230. Brief of Respondent at 6, State v. Strieby, 790 P.2d 98 (Utah Ct. App. 1990) (No. 890124-CA).

231. *Strieby*, 790 P.2d at 102.

232. Brown v. United States, 256 U.S. 335, 343 (1921).

Nor is the cessation theory relevant to establishing that Erlene did not reasonably perceive that danger was imminent. Chris was threatening to kill her, he had the present ability to do so, *and* he was advancing toward her when Erlene shot him. These facts would be sufficient to meet the imminence requirement under even the narrowest temporal proximity standard.

In contrast to the trial court and dissenting opinions, Judge Greenwood's majority opinion evinced a willingness to believe that Erlene's perceptions were reasonable. The court emphasized Erlene's claim that Chris's rage during the incident in question "had escalated beyond anything she had previously experienced with him."[233] She testified at trial that she "had never seen him like this."[234] Subsequently, in an interview with the Salt Lake Tribune, she said:

> I was never so scared for my life. . . . Most men who abuse don't abuse to the point the woman's going to die. She knows she's going to get a beating, but she doesn't feel she's going to die. . . .
>
> But that night I knew. There's a difference when you're fighting for your life. It's not an abuse situation anymore, it's murder. He's going to kill you and you know it. I wasn't guilty of anything but survival.[235]

Noting it had not doubted the veracity of this testimony, Judge Greenwood chided the trial court for nevertheless insisting on Erlene's unreasonableness by engaging in "pure speculation about a cessation of hostilities."[236] Not only did "[a]ll the evidence other than the blue cup" indicate that Erlene reasonably believed she was in imminent danger of death or serious bodily harm, but Erlene had in fact "already suffered serious injury."[237]

Judge Greenwood's opinion demonstrates a fair and unbiased application of black letter self-defense law to facts that cry out for its application. Perhaps *State v. Strieby* will encourage Utah lawyers in future battered women's homicide cases to dispel

---

233. *Strieby*, 790 P.2d at 101.
234. *See supra* text accompanying note 202.
235. Hobbs, *supra* note 196.
236. *Strieby*, 790 P.2d at 101.
237. *Id.*

anachronistic beliefs that might keep them from presenting their client's conduct as reasonable and justified.

IV. GUIDELINES

The law of self-defense must effect the valid policies that form the elements of self-defense without losing sight of the basic premise of justifiable homicide: "[W]here an individual cannot resort to the law in response to aggression, she may use reasonable force to protect herself from physical harm."[238] What is reasonable must be evaluated in light of all the circumstances. Care must be taken to avoid reliance on traditional biases that preclude women's equal treatment under the law. It is important that, where the facts so indicate, women's actions to protect themselves are seen as reasonable rather than excusable.

Juries should be allowed to consider evidence regarding the nature and extent of prior abuse, and patterns of escalating abuse. Such evidence points to the reasonableness of a woman's belief that her death was imminent. Empirical evidence regarding the dangers such abuse poses to the battered woman should be introduced to support the reasonableness of a woman's belief that she was defending her life. The brutal and life-threatening nature of the abuse should be stressed. Any change in the abuse is significant. It should be emphasized that the abused woman is in the best position to interpret the meaning of a change in her abuser's violence. She must also be allowed to show her reasons for believing his threats, and that given those reasons her actions were objectively reasonable.

Evidence tending to show a woman's appeals for help, and the extent to which she was trapped in her battering relationship by her circumstances, must also be considered. Though a woman should be legally justified in standing her ground, she nevertheless may be required to show why she did not exercise the alternative of leaving the abusive relationship. Any hardship or danger she encountered in trying to leave, or that she would have encountered had she left, should be considered. Only with such evidence before

238. *See* Faigman, *supra* note 134, at 624. In 1921, Justice Holmes pointed out that the law of self-defense had tended to "ossify into specific rules," not always fixed with regard to reason. Brown v. United States, 256 U.S. 335, 343 (1921).

it can the jury make an informed decision about the validity of her options. Her economic situation, number of dependent children, and the feasibility of alternative means of support are relevant as well. Furthermore, the number of times she attempted to leave but was retrieved by her abuser must be considered. She may not in fact have been "free to go." Finally, calls to the police and the police procedure in response to her calls should be brought out. The less able she was to elicit the assistance of the criminal justice system for her protection, the more reasonable appears her resort to self-protection.

V. CONCLUSION

Gender biases hamper battered women in their self-defense claims. While black letter self-defense law accommodates the factual circumstance many battered women's homicide cases present, self-defense is less likely to be argued and, as evidenced in *State v. Strieby*, less likely to be seen as reasonable. Yet, instead of focusing on gender bias in the *application* of long-standing self-defense principles, many battered women's advocates agitate for reform in the law itself. There may be instances in which reform is necessary. However, a disservice is done women if their pleas for equal treatment under the law are perceived as pleas for a major overhaul in the law to suit their specific needs. Moreover, pardoning and leniency does women a disservice. Often focusing on the passive and victimized aspects of women's experience, it compounds perceptions that women are not in fact justified in their actions, but rather that the law extends them favorable treatment. Instead, where a woman's actions to defend herself against an abusive spouse are objectively reasonable, in light of *all* the circumstances, the law must not excuse, but exonerate her.

PAIGE BIGELOW

Barnard v. Utah State Bar and Public Access to Private Entities Which Carry Out Governmental Functions: Is This Bar a Private Club?

Attorney Hugh W. Stroh, Jr., made the news when reporters discovered that he served his clients one week, then served a jail sentence the next, telling his clients that he was on vacation.[1] Stroh was convicted of attempting to persuade a police officer to lie while under oath.[2] Despite his conviction, Stroh avoided disbarment because the Washington State Bar ruled that the crime did not involve moral turpitude.[3] Stroh was allowed to stagger his thirty-day jail sentence over a four-month period so that he could serve his clients three weeks a month and tell his clients he was on vacation during the week he was in jail.[4]

Although some Americans with a "jaded view of the legal profession" may consider this story humorous, the American Bar Association is not laughing.[5] Attorneys are concerned about the legal profession's image.[6] Some commentators have suggested that secrecy in state bar disciplinary activities has created public distrust in the legal profession.[7] One remedy to this distrust is for state bars to voluntarily disclose their records to the public.[8] However, bars may choose not to open their records.

Another possible way to promote public confidence in the legal profession is for state legislatures to apply open records statutes[9] to state bars. *Barnard v. Utah State Bar*[10] arose from

---

1. Mary Collins, *Lawyers Have Chance to Scrutinize States That Turn Blind Eye to Corrupt Attorneys*, SALT LAKE TRIB., Jan. 28, 1992, at A9. The event occurred "a dozen years ago." *Id.*

2. *Id.*

3. *Id.*

4. *Id.*

5. *Id.*

6. *See id*; AMERICAN BAR ASS'N COMM'N ON EVALUATION OF DISCIPLINARY ENFORCEMENT, REPORT TO THE HOUSE OF DELEGATES (May 1991) [hereinafter "ABA REPORT"]; THE UTAH SUPREME COURT'S SPECIAL TASK FORCE ON THE MANAGEMENT AND REGULATION OF THE PRACTICE OF LAW, FINAL REPORT (Nov. 1991) [hereinafter "UTAH BAR REPORT"].

7. *See* ABA REPORT, *supra* note 6, at iv, 23, 24.

8. *See id.*

9. State disclosure laws consist of open records laws and open meetings laws. This Note focuses on open records laws. Some state disclosure statutes resemble the federal

the Utah State Bar's denial of a bar member's written request to obtain bar employee salary information. The Utah Supreme Court held that bar employee salaries were not subject to disclosure because the bar was not a state agency under Utah's disclosure acts.[11]

This Note suggests that open records laws should apply to state bars and other private entities, but only to the extent that they carry out functions delegated by governmental agencies.[12]

Freedom of Information Act, 5 U.S.C. § 552 (1988). Burt A. Braverman & Wesley R. Heppler, *A Practical Review of State Open Records Laws*, 49 GEO. WASH. L. REV. 720, 722 (1981). *See, e.g.*, Pasik v. State Bd. of Law Examiners, 451 N.Y.S.2d 570, 574 (Sup. Ct. 1982) (New York Freedom of Information Act patterned after federal version). However, many states have unique statutes. *See* ALA. CODE §§ 36-12-40 to -41 (1991); ALASKA STAT. §§ 09.25.110–.220 (Supp. 1991); ARIZ. REV. STAT. ANN. §§ 39-121 to -121.03 (1985); ARK. CODE ANN. §§ 25-19-101 to -107 (Michie 1987 & Supp. 1991); CAL. GOV'T CODE §§ 6250–6268 (West 1980 & Supp. 1992); COLO. REV. STAT. §§ 24-72-201 to -206 (1988 & Supp. 1991); CONN. GEN. STAT. §§ 1-15 to -21k (1991); DEL. CODE ANN. tit. 29, §§ 10001–10005, 10112 (1991); D.C. CODE ANN. §§ 1-1521 to -1529 (1987 & Supp. 1991); FLA. STAT. ANN. §§ 119.01–.16 (West 1982 & Supp. 1992); GA. CODE ANN. §§ 50-18-70 to -76 (1990 & Supp. 1991); HAW. REV. STAT. §§ 92-21, 92-50 to -52, 92F-11 to -42 (1988 & Supp. 1991); IDAHO CODE §§ 9-301 to -348 (1990 & Supp. 1991); ILL. ANN. STAT. ch. 116, paras. 201–211 (Smith-Hurd 1988 & Supp. 1991); IND. CODE ANN. §§ 5-14-3-1 to -10 (Burns 1987 & Supp. 1991); IOWA CODE ANN. §§ 22.1–.12 (West 1989 & Supp. 1991); KAN. STAT. ANN. §§ 45-215 to -225 (1986 & Supp. 1991); KY. REV. STAT. ANN. §§ 61.870–.884 (Michie/Bobbs-Merrill 1986 & Supp. 1990); LA. REV. STAT. ANN. §§ 44:1–:16, :31–:42 (West 1982 & Supp. 1992); ME. REV. STAT. ANN. tit. 1, §§ 401–410 (West 1989 & Supp. 1991); MD. CODE ANN., STATE GOV'T §§ 10-611 to -628 (1984 & Supp. 1991); MASS. ANN. LAWS ch. 4, § 7, cl. 26 (Law. Co-op. 1988 & Supp. 1991), ch. 66, § 10 (Law. Co-op. 1991); MICH. STAT. ANN. § 4.1801 (Callaghan 1985 & Supp. 1991); MINN. STAT. ANN. §§ 13.01–.90 (1988 & Supp. 1992); MISS. CODE ANN. §§ 25-61-1 to -17 (Supp. 1990); MO. ANN. STAT. §§ 109.180–.195 (Vernon 1966 & Supp. 1992); MONT. CODE ANN. §§ 2-6-101 to -111 (1991); NEB. REV. STAT. §§ 84-712 to -712.09 (1987 & Supp. 1991); NEV. REV. STAT. §§ 239.005–.330 (1991); N.H. REV. STAT. ANN. §§ 91-A:1–:8 (1990 & Supp. 1991); N.J. STAT. ANN. §§ 47:1A-1 to -4 (West 1989); N.M. STAT. ANN. §§ 14-2-1 to -3 (Michie 1988); N.Y. PUB. OFF. LAW §§ 84–90 (McKinney 1988 & Supp. 1992); N.C. GEN. STAT. §§ 132-1 to -9 (1991); N.D. CENT. CODE §§ 44-04-18 to -19.1 (1978 & Supp. 1991); OHIO REV. CODE ANN. §§ 149.43–.43.1 (Anderson 1990 & Supp. 1990); OKLA STAT. ANN. tit. 51, § 24A.1–.20 (West 1988 Supp. 1992); OR. REV. STAT. §§ 192.410–.595 (1991); PA. STAT. ANN. tit. 65, §§ 66.1–.4 (1959 & Supp. 1991); R.I. GEN. LAWS §§ 38-2-1 to -14 (1990 & Supp. 1991); S.C. CODE ANN. §§ 30-4-10 to -110 (Law. Co-op. 1990); S.D. CODIFIED LAWS ANN. §§ 1-27-1 to -19 (1985 & Supp. 1991); TENN. CODE ANN. §§ 10-7-503 to -507 (1987 & Supp. 1991); TEX. REV. CIV. STAT. ANN. art. 6252-17a (West Supp. 1992); UTAH CODE ANN. §§ 63-2-101 to -909 (Supp. 1992); VT. STAT. ANN. tit. 1, §§ 315–320 (1985 & Supp. 1991); VA. CODE ANN. §§ 2.1-340 to -346.1 (Michie 1987 & Supp. 1991); WASH. REV. CODE ANN. §§ 42.17.250–.450 (West 1991); W. VA. CODE §§ 29B-1-1 to -6 (1986); WIS. STAT. ANN. §§ 19.21–.39 (West 1986 & Supp. 1988); WYO. STAT. §§ 16-4-201 to -205 (1990 & Supp. 1991).

10. 804 P.2d 526 (Utah 1991).

11. *Id.* at 529.

12. Distinguishing between an entity's public and private activities is difficult. See

The Note recommends further that courts view records in camera[13] and issue protective orders, cloistering records when justice requires.[14] Section I discusses why state disclosure laws should apply to agencies that carry out the public's business. The section illustrates, however, the unique nature of state bars, which prevents some legislatures and courts from applying legislative enactments to state bars. Section II discusses the facts of *Barnard* and outlines the court's decision. Section III analyzes *Barnard* in light of decisions from the United States Supreme Court and other state courts. Section IV addresses whether Utah's open records provisions[15] affect the *Barnard* decision. Next, Section V discusses how the constitutional doctrine of separation of powers may prevent courts from applying state disclosure laws to state bars. Section VI then illustrates how some courts have granted public access to information concerning bar admission and disciplinary activities, and discusses whether a state's open records statute

---

Westchester Rockland Newspapers, Inc. v. Kimball, 430 N.Y.S.2d 574, 577-78 (1980). This challenge, however, should not prevent legislatures and courts from compelling private entities to disclose information regarding their public activities. One court has attempted to solve the problem by suggesting that records be viewed in camera to determine whether they are public or private. *See* Fritz v. Norflor Constr. Co., 386 So. 2d 899, 901 (Fla. Dist. Ct. App. 1980) (citing Shevin v. Byron, Harless, Schaffer, Reid & Assocs., 379 So. 2d 633 (Fla. 1980)).

The Utah State Bar carries out both private activities and governmentally delegated functions. For a list of Utah State Bar programs, see UTAH BAR REPORT, *supra* note 6, at exhibit B. Of the Utah State Bar's functions, the following arguably are governmentally delegated: conducting admissions, discipline, and licensing; regulating the unauthorized practice of law; and providing continuing legal education. The following bar activities are arguably private: sponsoring a VISA/Mastercard program for bar members; sponsoring Lexis programs in law offices; helping attorneys impaired by drugs; negotiating a group rate on insurance; engaging in lobbying on behalf of the legal profession; publishing the bar directory; providing a lawyer referral service; publishing the *Utah Bar Journal*; assisting in the delivery of legal services to the public; providing fee arbitration; maintaining a client security fund; regulating lawyer advertising; improving lawyer relations with courts and with judges; providing law-related education in schools; holding annual meetings; conducting public relations programs; providing law-related education via media; studying the legal needs of the elderly and of children; improving relations between professions; sponsoring law day activities and mock trials; sponsoring a young lawyers section; providing alternative dispute resolution facilities; conducting Tuesday night bar; and providing conference and meeting facilities. *Id.*

13. *See Fritz*, 386 So. 2d at 901.

14. *See* Carter v. Utah Power & Light Co., 800 P.2d 1095, 1100 (Utah 1990); UTAH CODE ANN. § 63-2-301(1)(f); David B. Dellenbach, Development, *Public Access to Judicial Records*, 1992 UTAH L. REV. 265; ABA REPORT, *supra* note 6, at 23-26.

15. *See* Government Records Access and Management Act ("GRAMA"), UTAH CODE ANN. §§ 63-2-201 to -308 (Supp. 1992).

should apply to a state bar's activities. Finally, Section VII considers the merits of the different statutes states might use to apply disclosure laws to bars and other private entities.

I. BACKGROUND

A. Purpose of State Open Records Statutes

One commentator has noted that "[t]he public's increasing disenchantment with public officials and the fear of political shenanigans or corruption has produced a desire to participate in the omnipresent government regulation of business and private lives."[16] State open records statutes enable citizens to participate in government by allowing them access to information about governmental activities.[17] However, many private agencies that carry out delegated public functions are not subject to state open records laws and thus can carry out the public's business in secret. State bars are good examples of private entities which perform some delegated public functions, as well as some private functions.

Given the probability of an increased interest in privatization, one commentator believes it is time to develop "more structured method[s] for dealing with delegation issues."[18] An important method to check private entities exercising governmental power is to require these agencies to disclose records of their publicly-funded activities.[19] For example, a Florida appeals court held

---

16. Laurie A. O'Connell, Comment, Common Cause v. Utah Public Service Commission—*The Applicability of Open-Meeting Legislation to Quasi-Judicial Bodies*, 1980 UTAH L. REV. 829, 829.

17. *See id.; see, e.g.*, IND. CODE § 5-14-3-1 (1988). The purpose of Indiana's disclosure statute is to ensure that "all persons are entitled to full and complete information regarding the affairs of government and the official acts of those who represent them as public officials and employees." *Id.; see also* Braverman & Heppler, *supra* note 9, at 722 (state open records laws provide access to "a nearly endless variety" of state and local government activities).

18. David M. Lawrence, *Private Exercise of Governmental Power*, 61 IND. L.J. 647, 647 (1986).

19. *See In re* Westchester Rockland Newspapers, Inc. v. Kimball, 430 N.Y.S.2d 574, 577 (1980). In *Westchester*, the court allowed public access to records of a lottery sponsored by a volunteer fire department. *Id.* at 575-78. In allowing access, the court relied on a statute which provided that "[a]s state and local government services increase and public problems become more sophisticated and complex and therefore harder to solve, and with the resultant increase in revenues and expenditures, it is incumbent upon the state and its localities to extend public accountability *wherever and whenever*

that the records of a private engineering firm were subject to public disclosure "insofar as [the firm] performed services for the city."[20]

Bar disciplinary activities and admission processes are arguably public functions because state supreme courts delegate these functions to state bars.[21] In other professions, including medicine,[22] a state agency is responsible for admissions and discipline. Therefore, the issue, discussed in Section VI, is whether bars should be subject to open records laws because they carry out these governmentally delegated functions.

B. Issues Involved in Applying State Open Records Statutes to State Bars

There are three main issues involved in applying open records laws to state bars. First, many state bars are classified as private entities while many state open records statutes apply only to public entities.[23] Before applying a state disclosure statute to any agency, courts first consider whether the legislature intended the statute to apply to the entity.[24] Courts must also consider

---

feasible." *Id.* at 576 (quoting N.Y. PUB. OFF. LAW § 84 (McKinney 1988)).

20. Fritz v. Norflor Constr. Co., 386 So. 2d 899, 901 (Fla. Dist. Ct. App. 1980); *see also* Schwartzman v. Merritt Island Volunteer Fire Dep't, 352 So. 2d 1230, 1231-32 (Fla. Dist. Ct. App. 1977) (holding nonprofit volunteer fire department a state agency, subject to records act), *cert. denied,* 358 So. 2d 132 (Fla. 1978); State *ex rel.* Bartow v. Public Employees Relations Comm'n, 341 So. 2d 1000, 1002-03 (Fla. Dist. Ct. App. 1976) (holding public employees commission subject to disclosure), *cert. denied,* 352 So. 2d 170 (Fla. 1977); Doe v. Sears, 263 S.E.2d 119, 121-22 (Ga. 1980) (holding city housing authority subject to disclosure laws). *But see* New York Post Corp. v. Moses, 176 N.E.2d 709, 710-12 (N.Y. 1961) (bridge and tunnel authority not subject to state disclosure laws).

21. *See* Barnard v. Utah State Bar, 804 P.2d 526, 529 (Utah 1991) (disciplinary functions governmentally delegated); Lawrence, *supra* note 18, at 667-68 (determining bar admission is governmental function).

22. *See* UTAH CODE ANN. §§ 58-12-26 to -43 (1990 & Supp. 1991).

23. *See, e.g., Barnard,* 804 P.2d at 529 (finding Utah State Bar a private entity and that records act did not apply). In determining whether an entity is subject to a state legislative enactment, courts consider the legislative history of the statute and whether the statute was intended to compel disclosure from the entity: "The cardinal rule of statutory interpretation is to determine the intent of the legislature in enacting the statute. The derivation of legislative intent begins with an analysis of the statute through examination of its plain language." Byron C. Keeling, Note, *Attempting to Keep the Tablets Undisclosed: Susceptibility of Private Entities to the Texas Open Records Act,* 41 BAYLOR L. REV. 203, 204-05 (1989).

24. Andrea G. Nedel, Annotation, *What Constitutes an Agency Subject to Application of State Freedom of Information Act,* 27 A.L.R.4th 742, 746 (1984). "The first step . . . is

whether the specific record sought is subject to the open records statute.[25] The second issue involved in applying state open records statutes to state bars is the doctrine of the separation of powers. Because courts, not legislatures, generally govern state bars, the separation of powers doctrine prevents legislatures from regulating the practice of law.[26] Thus, state courts may be the only governmental bodies empowered to compel disclosure from state bars. Third, there is a conflict between the need for privacy and self-regulation in the admissions and disciplinary processes and the need for public disclosure to engender public confidence.[27]

The *Barnard* decision discussed only the first of these three issues: whether or not the Utah State Bar is a state agency.[28] The court held that because the bar is not a state agency, Utah's open records provisions did not apply to it.[29] Therefore, the court did not consider whether the doctrine of separation of powers prevented the Utah Legislature from applying Utah's open records act to the bar.[30] However, Justice Stewart, writing for the majority, hinted that a separation of powers conflict might have existed

---

to decide whether the custodian of the record or report constitutes an agency subject to the provisions of the applicable state freedom of information law." *Id.*; *see also* Braverman & Heppler, *supra* note 9, at 730 (first step is to determine whether custodial agency is covered by the state disclosure law).

For application to government agencies, see City of Kenai v. Kenai Peninsula Newspapers, Inc., 642 P.2d 1316, 1323-24 (Alaska 1982) (state disclosure statute applied to applications for public office); Westinghouse Broadcasting Co. v. Sergeant-At-Arms, 375 N.E.2d 1205, 1208 (Mass. 1978) (legislature not an instrumentality enumerated in the disclosure act); Lodge v. Knowlton, 391 A.2d 893, 894-95 (N.H. 1978) (division of state police within department of safety a public agency); Dunlea v. Goldmark, 389 N.Y.S.2d 423, 425 (App. Div. 1976) (budget examiner's files on cable television commission subject to disclosure law); Barnett v. Long Island State Park Comm'n, 323 N.Y.S.2d 71, 73-74 (Sup. Ct. 1971) (state park subject to disclosure law).

25. Braverman & Heppler, *supra* note 9, at 732.

26. *See* Graham v. Washington State Bar Ass'n, 548 P.2d 310, 315-16 (Wash. 1976).

27. *See* ABA REPORT, *supra* note 6, at 86. A 1970 report by the American Bar Association's Clark Committee recommended that "it was more important to protect innocent lawyers [from harm to their reputations] than to notify the public of unproven allegations of the lawyer's misconduct . . . ." *Id.* Conversely, the 1991 McKay Committee found that little or no harm would come from public disclosure of allegations. *Id.* at 88.

28. *Barnard*, 804 P.2d at 527.

29. *Id.* at 529.

30. Section V discusses how the doctrine of separation of powers affects legislative attempts to regulate the practice of law.

had the court held that the legislature intended the statute to apply to the bar.[31]

C. Broad and Flexible Versus Narrow and Rigid Statutes

The unique nature of state bars presents a formidable challenge to legislatures and courts attempting to apply legislative enactments to bars. Courts usually will not consider state bars as public agencies under narrow, rigid statutes that apply only to entities performing traditional governmental functions.

For example, in *In re Rhode Island Bar Association*,[32] the Rhode Island Supreme Court held that the Rhode Island Bar was not a state agency under the state's administrative procedure act.[33] The court reasoned that the bar did not qualify under the statute which defined "agency" as a "state board, commission, department, or officer, other than the legislature or the courts, authorized by law to make rules or to determine contested cases."[34] The bar was "not empowered to determine contested cases or to make rules" without the approval of the Rhode Island Supreme Court. Accordingly, the court concluded that the bar was not a state agency.[35]

The *Rhode Island Bar* case illustrates how bars do not fit neatly within narrow, rigid state statutes which apply only to entities that typically carry out most state functions, such as boards, commissions, and departments. However, broader, more flexible statutes can be applied successfully to state bars: "A combination of a liberal 'public record' definition and a broad 'public agency' definition based upon public funding may bring normally excluded nongovernmental entities, such as consultants

---

31. *See Barnard*, 804 P.2d at 530. The *Barnard* court explained:

> The Utah Constitution assigns to this court the power to "govern the practice of law." UTAH CONST. art. VIII, § 4. We need not, and therefore do not decide whether that grant ousts the Legislature from all control over the Bar or whether the Records Act and Writings Act would be unconstitutional if applied to the Bar.

Id.

32. 374 A.2d 802 (R.I. 1977).
33. *Id.* at 803.
34. *Id.* (citing R.I. GEN. LAWS § 42-35-1(a) (1969)).
35. *Id.*

and quasi-public corporations, within the reach of a state's open records law."[36]

For example, in *Connecticut Bar Examining Committee v. Freedom of Information Commission,*[37] the Connecticut Supreme Court, applying Connecticut's broad statute,[38] held that the bar examining committee was subject to the state freedom of information statute. Connecticut's disclosure law expressly applied to judicial bodies or committees, but "'only in respect to [their] . . . administrative functions.'"[39] The court thus ruled in favor of access to records relating solely to the bar examining committee's administrative functions.[40] Finally, the court remanded the case, directing the trial court to determine whether public access to the administrative records would interfere with the committee's judicial functions.[41]

Connecticut's statute is broad and flexible enough to encompass bar activities.[42] The statute, however, limits disclosure to the administrative functions of the bar examining committee,[43] thereby balancing the interests of privacy against the public's right to know how the admission process operates.

Florida's disclosure statute is another example of flexibility. It applies to any "public or private agency . . . acting on behalf of any public agency."[44] Under Florida's statute, a towing company that had a contract with a city to "remove vehicles from the streets . . . only as directed by an authorized representative of the City

---

36. Braverman & Heppler, *supra* note 9, at 731.

37. 550 A.2d 633 (Conn. 1988).

38. "'Public agency' . . . means any executive, administrative, or legislative office . . . and includes any committee of any such office, subdivision, agency, department, institution, bureau, board, commission, authority or official, and also includes any judicial office, official or body or committee thereof but only in respect to its or their administrative functions." CONN. GEN. STAT. § 1-18a(a) (1991).

39. *Connecticut Bar,* 550 A.2d at 635 (quoting CONN. GEN. STAT. § 1-18a (1991)).

40. *Id.* at 636.

41. *Id.*

42. *See id.*

43. *Id.*

44. FLA. STAT. ANN. § 119.011(2) (West 1982). In Florida, "agency" is defined as "any state, county, district, authority, or municipal officer . . . and any other public or private agency, . . . or business entity acting on behalf of any public agency." *Id.* Public records include "all documents, [and] papers . . . made or received pursuant to law or ordinance or in connection with the transaction of official business by any agency." *Id.* § 119.011(1).

police department" was held to be carrying out a public function and was considered to have possession of public records.[45]

The Florida statute reflects an attempt to apply disclosure laws to all entities to the extent they are acting on behalf of the government, while protecting private activities. The Connecticut and the Florida statutes have enough flexibility to reach private enterprises exercising governmental power, in order to provide for checks and balances, and to engender public trust.

D. Utah's Records and Writings Acts

Before the Government Records Access and Management Act[46] ("GRAMA") became effective on April 1, 1992, Utah's open records acts were rigid statutes. They applied only to a limited list of state entities and contained no general or flexible provisions which might apply to an entity not fitting neatly within the rigid categories. The statutory scheme consisted of two acts under which Utah courts had allowed public access to government records: the Public and Private Writings Act[47] ("Writings Act") and the Archives and Records Services and Information Practices Act[48] ("Records Act").[49] The Writings Act granted the public the right of access to "public writings."[50] The definition of "public writings"

---

45. Fox v. News-Press Pub. Co., 545 So. 2d 941, 942-43 (Fla. Dist. Ct. App. 1989).

46. UTAH CODE ANN. §§ 63-2-101 to -308 (Supp. 1991).

47. UTAH CODE ANN. §§ 78-26-1 to -8 (1987) (§§ 78-26-1 to -3 *repealed by* Government Records Access and Management Act, ch. 259, 1991 Utah Laws 1007, 1043-44 (effective April 1, 1992)). For comparable provisions, see Government Records Access and Management Act, ch. 259, 1991 Utah Laws 1007, 1012 (codified as amended at UTAH CODE ANN. §§ 63-2-201 to -308 (Supp. 1991) (effective April 1, 1992)). *See generally* Christian J. Rowley, Development, *Government Records Access and Management Act*, 1992 UTAH L. REV. 375 (discussing repeal of key provisions of Disclosure Act and enactment of Government Records Access and Management Act).

48. UTAH CODE ANN. §§ 63-2-59 to -89 (1989) (§§ 63-2-59 to -71, 63-2-73, 63-2-75 to -80, 63-2-84, 63-2-85.1 to -89 *repealed by* Government Records Access and Management Act, ch. 259, 1991 Utah Laws 1007, 1043-44 (effective April 1, 1992)). For comparable provisions, see Government Records Access and Management Act, ch. 259, 1991 Utah Laws 1007, 1012-20 (codified as amended at UTAH CODE ANN. §§ 63-2-201 to -308 (Supp. 1991)).

49. *See* KUTV, Inc. v. Utah State Bd. of Educ., 689 P.2d 1357, 1360-61 (Utah 1984) (student questionnaires subject to disclosure under Writings Act); Redding v. Brady, 606 P.2d 1193, 1195 (Utah 1980) (granting access to salaries of state college employees under both Writings Act and Records Act).

50. UTAH CODE ANN. § 78-26-2 (1987) (repealed 1992).

contained in the Writings Act included "public records."[51] Therefore, Utah courts looked to the definition of "public records" in Utah's Records Act to determine which records were subject to disclosure under the Writings Act."[52]

The Records Act defined "public records" as:

> all books, papers, letters, documents, maps, plans, photographs, sound recordings, management information systems, or other documentary materials, regardless of physical form or characteristics, made or received, and retained by any state public office under state law or in connection with the transaction of public business by the public offices, agencies, and institutions of the state and its counties, municipalities, and other political subdivisions.[53]

Since the Records Act defined public records as those records "retained by any state public office," The Barnard court also looked to the Records Act's definition of "public office" to determine if the bar's records were public.[54] "Public offices" included agencies of the state.[55] A "state agency," in turn, was a "department, division, board, bureau, commission, council, institution, authority, or other unit, however designated, of the state."[56] Thus, whether the Writings Act compelled disclosure of a record depended upon whether the entity holding the record was considered a department, division, board, commission, or institution, "of the state," and hence a "state agency."[57] The dispositive issue in *Barnard* was whether the Utah State Bar was a "state agency" under the Records Act.[58]

---

51. "Public Writings are divided into four classes: (1) Laws. (2) Judicial records. (3) Other official documents. (4) Public records . . . ." *Id.* § 78-26-1.

52. *See* Barnard v. Utah State Bar, 804 P.2d 526, 527 (Utah 1991).

53. UTAH CODE ANN. § 63-2-61(1) (1989) (repealed 1992).

54. *See Barnard*, 804 P.2d at 527.

55. UTAH CODE ANN. § 63-2-61(3) (1987) (repealed 1992).

56. *Id.* § 63-2-61(2).

57. *See Barnard*, 804 P.2d at 527. The Utah State Bar conceded that if it was considered a state agency under the Records Act, the Writings Act would also apply to the bar, mandating disclosure of bar records. *Id.* at 527 n.1.

58. *Id.* at 527.

II. THE CASE: *Barnard v. Utah State Bar*

In 1987, Brian M. Barnard, a member of the Utah State Bar, sent a letter to the bar's executive director requesting information regarding the salaries and benefits paid to its employees.[59] In response, the executive director disclosed the salary ranges for categories of bar employees and described the fringe benefits of the bar staff. Not satisfied, Barnard wrote another letter requesting more specific information. The executive director refused any further disclosure. Barnard then filed suit against the executive director and the bar, requesting a decree that the bar provide him with the specific salary information.

Barnard argued that the bar was a state agency and was therefore required to disclose the information pursuant to the Writings Act and the Records Act. Barnard also sought attorney fees and exemplary damages. The Third District Court ruled in favor of disclosure of the salary information, but did not award attorney fees or exemplary damages.

On appeal, the bar argued that it was not a state agency and therefore was not required to comply with the Records and Writings Acts. Instead, the bar claimed to be a private organization performing public service functions.[60] The Utah Supreme Court agreed and reversed the district court, holding that the bar was not a state agency subject to the disclosure provisions of the Records and Writings Acts.[61] In deciding that the bar was not a state agency, the court considered[62] the factors articulated by the United States Supreme Court in *Keller v. State Bar of California*:[63] the nature, purpose, and functions of the bar,[64] and the bar's statutory history.[65]

---

59. *Id.* at 526. All facts discussed here are from the court's opinion. *Id.* at 526-27.

60. The Utah State Bar assists the supreme court in regulatory functions, including administration of the bar examination and recommending disciplinary actions. *Id.* at 529.

61. *Id.* at 529-30. The court's holding was limited to the conclusion that the bar was not a state agency under the state Records and Writings Acts. *See id.* The court did not examine whether the Utah State Bar is a state agency under other acts, such as the federal Civil Rights Act, 42 U.S.C. § 1983 (1981). *See Barnard*, 804 P.2d at 527. The court also did not decide whether the plaintiff could have petitioned the court, under its rulemaking authority, to require the Utah State Bar to disclose the information. *See id.*

62. *Barnard*, 804 P.2d at 527.

63. 110 S. Ct. 2228, 2235-36 (1990).

64. *Barnard*, 804 P.2d at 529-30.

65. *Id.* at 528.

A. *The Test Applied in* Keller v. State Bar of California

Keller is a leading case addressing the public/private character of state bars.[66] In *Keller*, the California State Bar attempted to establish "governmental agency" status so that it could take substantive positions on political issues under the "government speech doctrine."[67] The United States Supreme Court rejected the argument, holding that the California State Bar was not a governmental entity. The *Keller* court identified two key elements as distinguishing the California Bar from other governmental agencies: private funding, and the advisory nature of the bar services.[68] The *Keller* Court, comparing the California Bar to an employee union,[69] held that the state's interest in regulating the legal profession justifies the compelled association of an integrated bar, but the bar may not use compulsory bar dues to fund activities which are not germane to the goals of all bar members.[70]

Like the state bar in California, the Utah State Bar administers qualifying examinations and recommends disciplinary proceedings against attorneys, but the Utah Supreme Court retains final decision-making authority over these matters.[71] Furthermore, the Utah Supreme Court retains power over disbarment, suspension, public reprimand, and it approves[72] the

66. The decision in *Keller* caused state bars across the country to reevaluate their lobbying activities. Bruce Hamilton, Keller *Provides a Clean Bill of Health*, 27 ARIZ. ATT'Y, Dec. 1990, at 30; *see* James K. Robinson, *Meeting* Keller's *Challenge to the Future of Michigan's Integrated Bar*, 70 MICH. B.J., June 1991, at 516.

67. *Keller*, 110 S. Ct. at 2234.

68. *See id.* at 2234-35.

69. *Id.* at 2235-38 (citing Abood v. Detroit Bd. of Educ., 431 U.S. 209, 235-36 (1977)). In Abood v. Detroit Board of Education, 431 U.S. 209 (1941), the United States Supreme Court held that a union could use union dues to fund political activities not germane to collective bargaining only when the employees did not object to those political activities. *Id.* at 235-36.

70. *Keller*, 110 S. Ct. at 2235-37.

71. *See Barnard*, 804 P.2d at 529.

72. Other courts have recognized that entities acting without the need of governmental approval are private entities. For example, New York's disclosure statute defines "agency" as "any state or municipal department, board, bureau, division, commission, committee, public authority, public corporation, council, office or other governmental entity performing a governmental or proprietary function for the state or any one or more municipalities thereof." N.Y. PUB. OFF. LAW § 86 (McKinney 1988). A New York appellate court held that the Buffalo Economic Development Corporation was not a public agency subject to the state's freedom of information act, despite the fact that the

rules governing admission to the bar.[73] Additionally, the Utah State Bar's advisory/regulatory services do not make it a public agency.[74] Finally, the bar is funded by its members, not public funds.[75]

In addition to applying *Keller*'s analysis, the Utah Supreme Court weighed the Utah State Bar's private functions against its public functions. In determining that the bar's public functions did not transform it into a public entity, the court considered the various nongovernmental attributes of the bar.[76] The bar can "sue and be sued," "[i]t owns real property in its own name," and it pays taxes on that property.[77] It is self-governed by uncompensated bar commissioners who are elected by bar members.[78] The bar is funded completely by members' dues and receives no public funds.[79] The bar hires its own counsel to defend lawsuits.[80] Finally, the bar's budget is not subject to legislative approval.[81] After applying the *Keller* test and balancing the bar's public functions against its private functions, the *Barnard* Court concluded that the Utah State Bar was not a public agency.[82]

B. The Nature, Function, and Purpose of the Bar

The *Barnard* court found that the nature of the Utah State Bar made it difficult to determine whether the bar was a state or

---

agency received public funds and received free rent in an office in Buffalo City Hall. *See* Buffalo News Inc. v. Buffalo Enter. Dev. Corp., 561 N.Y.S.2d 406, 407 (Sup. Ct. 1990). The court reasoned that the agency cooperated with local government, but acted independently, without need of governmental approval to function. The court further recognized that the government was not involved in the core planning and execution of the agency. *Id.* "The corporation's decisions with regard to lending and foreclosure are independent, not subject to review of any governmental council." *Id.; see also* A.S. Abell Publishing Co. v. Mezzanote, 464 A.2d 1068, 1072 (Md. 1983) (in determining whether statutorily established entity is a state agency, court should consider many factors, including degree of control exercised by state over entity).

73. *See Barnard*, 804 P.2d at 529.
74. *Id.* at 529-30.
75. *Id.*
76. *Id.* at 530.
77. *Id.*
78. *Id.*
79. *Id.*
80. *Id.*
81. *Id.*
82. *Id.* at 529-30.

private agency, because the bar is an organization *sui generis*,[83] meaning "[o]f its own kind."[84] State bars carry out two main functions: serving the public at large, and serving bar members.

On the one hand, bars contribute to their communities, not by providing typical governmental services but by improving the quality of legal services within a state. Bars do not exist to "participate in the general government of the State, but to provide specialized professional advice to those with the ultimate responsibility of governing the legal profession."[85] The Utah State Bar performs additional community service functions, including Tuesday night bar, law-day events, and public education.[86] State bars are essential to states because they assist courts in regulating the practice of law without relying on taxpayer funding.

On the other hand, state bars improve attorneys' ability to practice law. By disseminating knowledge of legal principles, bars "enhance professional competence."[87] Attorneys benefit from a host of bar services, ranging from continuing legal education to programs for lawyers impaired by alcohol and drugs.[88] Of course, these services also improve the practice of law within communities.

The *Barnard* court did not expressly state whether the nature of the Utah State Bar was more private than public. The unique features of the bar make it difficult to determine whether it falls within any of the categories used in Utah's Records Act to define a state agency.

C. The History of The Utah State Bar

The *Barnard* court found that the history of the Utah State Bar was not dispositive on the issue of whether the bar is a private or public agency. Before 1931, the Utah Legislature and Utah courts provided for discipline of attorneys and for admission to practice law, while the bar was a voluntary organization of

---

83. *Id.* at 529. Some might argue that the Utah State Bar is no longer *sui generis* because it is now an incorporated entity.

84. BLACK'S LAW DICTIONARY 1434 (6th ed. 1990).

85. Keller v. State Bar of California, 110 S. Ct. 2228, 2235 (1990).

86. UTAH BAR REPORT, *supra* note 6, at exh. B.

87. Barnard v. Utah State Bar, 804 P.2d 526, 529 (Utah 1991).

88. *See* UTAH BAR REPORT, *supra* note 6, at exh. B.

attorneys.[89] In 1931, the legislature began to require that all practicing attorneys join the bar and pay a license fee to a fund which was used by the bar.[90] The legislature also empowered the bar to recommend the admission of applicants and discipline of attorneys to the Utah Supreme Court.[91] Notwithstanding this statutory recognition of the bar, it is the Utah Supreme Court that has had the inherent power to "regulate the admission and discipline of attorneys . . . from the beginning."[92]

In 1981, the Utah Supreme Court integrated[93] the bar under its own "inherent power derived from the historic and fundamental relationship between attorneys at law and the courts and the doctrine of separation of powers."[94] In 1985 the Utah Legislature expanded the court's powers to regulate the practice of law by amending the Utah Constitution.[95] The amendment provides that "[t]he supreme court by rule shall govern the practice of law, including admission to practice law and the conduct and discipline of persons admitted to practice law."[96]

The history of the Utah State Bar reveals that a separation of powers conflict may prevent the Utah Legislature from applying any open records legislation to the state bar. However, the history does not reveal whether the Utah Legislature *intended* the bar to be a state agency. Thus, the basis for the *Barnard* court's ruling that the Utah State Bar was not a state agency rested solely on the *Keller* test and the conclusion that the bar's private functions were not outweighed by its public functions.

---

89. *Barnard*, 804 P.2d at 528.

90. *Id.*

91. *Id.*

92. *Id.* (citing *In re* McCune, 717 P.2d 701, 704-05 (Utah 1986)).

93. The rules for integration provide that "[a]ll persons now or hereafter licensed in this State to engage in the practice of law shall be members of the Utah State Bar . . . ." RULES FOR INTEGRATION AND MANAGEMENT OF THE UTAH STATE BAR, Rule A(1). There has been a great deal of controversy and comment about the compulsory nature of integrated bars. *See* Charles W. Sorenson, Jr., *The Integrated Bar and the Freedom of Nonassociation—Continuing Siege*, 63 NEB. L. REV. 30 (1983); Christopher Yost, Note, *Belly Up to the Bar: Your Bar Tab Is Compelled Membership and Mandatory Fees*, 20 PAC. L.J. 1281, 1281 (1989).

94. *Barnard*, 804 P.2d at 528.

95. *Id.*

96. UTAH CONST. art. VIII, § 4.

III. ANALYSIS OF THE *Barnard* COURT'S DECISION

The Utah Supreme Court accurately applied the *Keller* test and correctly determined that under the Writings Act and the Records Act, the Utah State Bar was a private entity. Furthermore, the *Barnard* decision is in accord with the test applied in many state courts to determine whether an agency is private or public.

A. "Totality-of-Circumstances" Test

In deciding whether a corporation or other organization is a public agency and thus subject to a state's disclosure statute, some courts consider a totality-of-factors test.[97] Among the factors considered under this test are: (1) whether the government created the corporation; (2) whether the government funds the organization; (3) the goals and purposes of the corporation; (4) the corporation's ownership; and (5) the interdependence between the corporation and a public entity.[98] Considering these factors, a Florida court held that Florida's broad disclosure laws applied to a not-for-profit hospital corporation which received county funds, acted on behalf of the county hospital board, and assisted in the hospital board's activities.[99]

1. Government Creation / Recognition

The Utah Supreme Court does not give great deference to the first prong of the totality-of-factors test. In *Utah Technology Finance Corp. v. Wilkinson,*[100] the court held that a state development agency, officially created by statute and endowed with

97. *See* Sarasota Herald-Tribune v. Community Health Corp., 582 So. 2d 730, 733 (Fla. Dist. Ct. App. 1991) (citing Fox v. News Publishing Press Co., 545 So. 2d 941, 943 (Fla. Dist. Ct. App. 1989)).

98. *See id.* at 734; *see also* Connecticut Humane Soc'y v. Freedom of Info. Comm'n, 591 A.2d 395, 399 (Conn. 1991) (humane society not a public agency subject to state disclosure laws, despite being chartered and statutorily authorized by government, because it received no public funds and its activities were not subject to governmental review).

99. *See Sarasota,* 582 So. 2d at 734.

100. 723 P.2d 406 (Utah 1986).

public funds, was not a state agency.[101] The *Barnard* court cited *Utah Technology* for the proposition that "the mere fact that an organization is created or officially recognized by statute does not make it a state agency."[102] This conclusion is subject to question, however, because the Utah Supreme Court has made bar membership mandatory,[103] a fact that weighs in favor of finding that the Utah State Bar is a state agency.

2. *Government Funding*

Although government funding does not necessarily compel a finding that an organization is a public agency,[104] courts consider public funding to be a key factor in making the determination.[105]

---

101. *Id.* at 414-15. In 1953, the Utah Legislature created the Utah Technology Finance Corporation ("UTFC"). *Id.* at 408. The purpose of the corporation was to assist and develop high-tech businesses in the state. *Id.* The Utah State Treasurer was the custodian of the UTFC's funds and the UTFC was required to report to the Governor. *Id.* at 415. However, the Utah Supreme Court held that the UTFC was not an agency of the executive branch, but was an independent, nonprofit corporation. *Id.*

102. Barnard v. Utah State Bar, 804 P.2d 526, 529 (Utah 1991).

103. *Id.* at 528.

104. Some state statutes provide that entities are public only when a certain percentage of their budgets come from public funds. For example, the Michigan Freedom of Information Act, which applies to all public bodies, defines "public bodies" as including any "body which is created by state or local authority or which is primarily funded by or through state or local authority." MICH. STAT. ANN. § 4.1801(2)(b)(iv) (Callaghan 1985). Under this statute, a Michigan appellate court found a nonprofit foster-care corporation to be a nonpublic body that was not subject to the disclosure provisions of Michigan's Freedom of Information Act. The court based its opinion on the fact that the corporation received less than half of its operating funds from the government. *See* Kubick v. Child & Family Servs., 429 N.W.2d 881, 883 (Mich. Ct. App. 1988). Despite the fact that the government was the largest single source of funding for the corporation, the court held that the government was not the primary source of funding because the government provided less than 50% of its budget. *Id.* The Michigan statute fails to address the public's right to know what the foster-care corporation does with public funds, no matter what portion of the agency's budget derives from public sources. Additionally, the Michigan statute rigidly defines "public agency," and private organizations usually do not fit neatly within rigid classifications. The statute thus allows private agencies to use public funds secretly.

In contrast, the Arkansas open records statute broadly applies to any "agency wholly or partially supported by public funds or expending public funds." ARK. CODE ANN. § 25-19-103 (Michie 1987); *see also* Braverman & Heppler, *supra* note 9, at 730 (discussing breadth of Arkansas statute).

105. Despite the Michigan court's holding in *Kubick* that the foster-care corporation was not a public agency, receipt of public funds is often a determining factor in deciding whether an entity is a public agency subject to state disclosure laws. *See* Board of Trustees of Woodstock Academy v. Freedom of Info. Comm'n, 436 A.2d 266, 271 (Conn. 1980). "Since Woodstock Academy performs a basic governmental function in providing

Often, whether receiving public funds classifies an entity as a public agency depends on the specific state statute. For example, the Indianapolis Convention and Visitors Association was held to be a public entity under a statute which defined a public entity as any provider "[s]upported in whole or in part by appropriations or public funds or by taxation."[106]

Although the Utah Records Act did not require a state agency to be supported by public funds, the fact that the Utah State bar does not receive public funds[107] supports the argument that the bar is a private agency. Moreover, the fact that the Bar is self-supporting is probably the most important reason not to classify it as a public entity for all purposes.

Applying disclosure laws to institutions receiving public funds achieves the important public policy goal of providing checks and balances to the agency's control over those funds. For example, in *Redding v. Brady*,[108] the Utah Supreme Court granted a student newspapers editor's request to obtain Weber State College employee salary information.[109] In requiring disclosure, the court reasoned that the public's right to know what the state-funded college does with taxpayer money outweighs the employees' interest in privacy.[110] The court stated that "by accepting employment at the college its employees are not merely private citizens, but become public servants in whose conduct and in whose salary the public has a legitimate interest."[111]

Based upon the *Redding* court's logic, members of the public have no right to know a bar employee's salary because bar employee salaries are paid from funds generated from membership dues rather than public funds.[112] The privacy interest of bar

---

public education at a secondary school level, is nearly entirely . . . publicly financed, has its operations examined and certified by the state board of education so as to be eligible for reimbursement for tuition fees by local towns and for other statutory benefits, and is an entity created by statute for the sole purpose of maintaining a public school . . . , it must be considered a public agency . . . ." *Id.*

106. Indianapolis Convention & Visitors Ass'n v. Indianapolis Newspapers, Inc., 577 N.E.2d 208, 212 (Ind. 1991); IND. CODE ANN. § 5-11-1-16(e) (2) (Burns 1987).

107. *See Barnard*, 804 P.2d at 530.

108. 606 P.2d 1193 (Utah 1980).

109. *Id.* at 1194-97.

110. *Id.* at 1196-97.

111. *Id.* at 1196.

112. *Barnard*, 804 P.2d at 530. Moreover, Utah State Bar "[e]mployees are not paid by the state and are not entitled to any benefits given state employees." *Id.*

employees weighs against disclosure because there is no counterweight of the public interest in knowing what the government does with taxpayer dollars.

3. *Goals and Purposes of the Utah State Bar*

The goals and purposes of the Utah State Bar do not necessarily support a finding that it is a state agency. The bar has many goals and purposes,[113] some that only benefit attorneys, and some that serve society as a whole. The Utah Supreme Court has delegated some aspects of attorney discipline and admission to the Utah State Bar.[114] However, the bar also participates in purely private actions. For example, the bar negotiates group rates on insurance as a private service to attorneys.[115]

4. *Ownership of the Utah State Bar*

The State has no ownership interest in the Utah State Bar. According to the *Barnard* court, the Utah State Bar "owns real property in its own name, and the State has no interest therein."[116] Furthermore, the bar pays taxes on its property,[117] and

---

113. Some primary purposes of the Utah State Bar, as stated in the Bar's articles of incorporation, are

> to advance the administration of justice according to law, to aid the courts in carrying on the administration of justice, to provide for and regulate the admission of persons seeking to engage in the practice of law, to provide for the regulation and discipline of persons engaged in the practice of law, to foster and maintain on the part of those engaged in the practice of law high ideals of integrity, learning, competence and public service, and high standards of conduct, to provide a forum for the discussion of subjects pertaining to the practice of law, the science of jurisprudence, and law reform, to carry on a continuing program of legal research in technical fields of substantive law, practice and procedure, and to make reports and recommendations thereon, to encourage practices that will advance and improve the honor and dignity of the legal profession; and to the end that the responsibility of the legal profession and the individual members thereof may be more effectively and efficiently discharged in the public interest . . . .

Articles of Incorporation of Utah State Bar (June 21, 1991) (Office of the Secretary of State) [hereinafter Articles of Incorporation of Utah State Bar].

114. *Barnard*, 804 P.2d at 529. Having no final decision-making authority, the Utah State Bar may only recommend to the Utah Supreme Court what it considers appropriate action. *Id.*

115. UTAH BAR REPORT, *supra* note 6, at exh. B.

116. *Barnard*, 804 P.2d at 530.

is incorporated as a nonprofit organization.[118] The state's lack of ownership interest in the bar and its assets suggests that the bar is a private agency.

5. Interdependence Between the Utah State Bar and the Utah Supreme Court

The interdependence prong of the totality-of-factors test favors holding that the Utah State Bar is a public agency. The bar aids the Utah Supreme Court in advisory and administrative services.[119] The court depends on the bar to perform investigations into the moral conduct of bar applicants and to recommend proposed disciplinary actions.[120] Furthermore, the state requires that all attorneys belong to the bar.[121] Without this mandate, some attorneys would avoid paying bar dues by refusing to join. Moreover, without the mandatory dues, the bar would not be able to carry out its administrative and public service functions.[122]

6. Balancing the Factors

One could argue that the court might have gone either way under a totality-of-factors analysis. The interdependence and creation factors lean in favor of defining the Utah State Bar as a state agency, while the ownership and funding prongs lean the other way. The goals and purposes factor is arguably neutral. The fact that the bar pays 100% of its own budget strongly suggests that it is a private agency. On balance, and particularly in light of this latter fact, the court correctly decided that the bar is a private organization.

---

117. *Id.*

118. Articles of Incorporation of Utah State Bar, *supra* note 113, at art. VII. The articles of incorporation do not state that the members of the Utah State Bar, any other person, or any other entity "owns" the bar. The articles provide that, upon dissolution, the assets of the bar shall be disposed of by trustees in a manner consistent with its purpose. *Id.* at art. XI.

119. *Barnard,* 804 P.2d at 529-30.

120. *Id.* at 529.

121. *Id.* at 528.

122. UTAH BAR REPORT, *supra* note 6, at vi-vii.

B. Additional Factors Supporting the Court's Decision

The specific facts of *Barnard* do not weigh in favor of public access as would a case involving attorney discipline or access to information regarding the bar admissions process.[123] Barnard may be entitled to receive the salary information he sought, but he erred in bringing a claim under Utah's open records statutes, which grant access to all members of the public. Significantly, the Utah Supreme Court emphasized that it did not decide whether Barnard was entitled to the information by petitioning the court promulgate a rule requiring the Utah State Bar to disclose such information.[124] This dictum suggests that the court is inviting bar members to request a new rule allowing bar members access to bar employee salary information. This dictum also suggests that the court prefers to solve bar member disputes through the administrative process rather than through litigation. The court's dictum is apparently rooted in the correct assumption that a bar member's ability to receive information about bar employees should be based on her membership in the bar, not on her membership in the public at large.

Moreover, even if the Utah Legislature intended the open records laws to apply to the Utah State Bar, these statutes should not apply to private activities carried out by the bar. Just as the public arguably has no right of access to contracts in which a towing agency engages with a nongovernmental agency,[125] the public has no right of access to computer training or credit card programs offered for attorneys[126] when neither of these activities are sponsored by public funds. A rigid application of public disclosure laws to *all* activities of the bar would infringe on privacy interests.

---

123. *See infra* Section VI (analyzing whether state bars should disclose disciplinary records and bar admission proceedings). Although the plaintiff in *Barnard* was a bar member, the issue addressed in that case was whether Utah's public disclosure laws applied to the Utah State Bar (allowing members of the general public access to the salary information), not whether a bar member could gain such access. *Barnard*, 804 P.2d at 526-27.

124. *Barnard*, 804 P.2d at 527.

125. *Cf. supra* notes 44-45 and accompanying text (company under contract with city possessed public records only insofar as records related to city).

126. UTAH BAR REPORT, *supra* note 6, at exh. B.

Although privacy rights should protect the bar's salary records, the public should have access to some disciplinary activities and administrative activities in the admissions process.[127] The *Barnard* court's broad holding illustrates how Utah's rigid Records Act did not allow for public access to the governmentally-delegated functions carried out by the Utah State Bar or by other private entities in Utah. The problem illustrated in *Barnard* is that Utah's Records Act did not apply to *any* bar activities. A broader statute would permit courts to consider whether, in the particular activity which the bar undertakes, the bar performs a function specifically delegated by the government. As the Washington Supreme Court noted in *Graham v. Washington State Bar Association*,[128] "the meaning of the term 'agency' depends on its context."[129] A state bar can be a public entity for some purposes and a private entity for others.[130]

The remainder of this Note discusses: (1) whether GRAMA[131] allows public access to some Utah State Bar activities; (2) whether the doctrine of separation of powers prevents the Utah Legislature from legislating with regard to the Utah State Bar; and (3) whether state disclosure laws should apply to the public functions carried out by the Utah State Bar.

IV. GRAMA'S EFFECT ON *Barnard*

The 1991 version of GRAMA does not appear to affect the *Barnard* court's holding. In 1991, the Utah Legislature enacted GRAMA, amending key provisions of the Records Act and Writings Act.[132] Under GRAMA, records possessed by "governmental" entities are open to public access unless the records are private, confidential, protected, or otherwise restricted by statute.[133]

---

127. *See infra* Section VI (arguing that public should have access to some disciplinary and admissions proceedings).

128. 548 P.2d 310 (Wash. 1976).

129. *Id.* at 312.

130. *See* Goldfarb v. Virginia State Bar, 421 U.S. 773, 791 (1975) (holding bar was state agency by law, but not protected from private antitrust laws regulating private activity) (discussed in John E. Lopatka, *The State of "State Action" Antitrust Immunity: A Progress Report*, 46 LA. L. REV. 941, 970 n.1501 (1986)).

131. UTAH CODE ANN. §§ 63-2-101 to -909 (Supp. 1992).

132. *See supra* note 46-48 and accompanying text.

133. UTAH CODE ANN. §§ 63-2-201(3)(b), -301 to -305 (Supp. 1992).

The 1991 edition reaches further by compelling disclosure from "[entities] in the executive branch that [are] publicly funded or that [are] established by the government to carry out the public's business."[134] The statute significantly broadened the scope of Utah's disclosure laws and transformed private entities into governmental entities, subject to disclosure, if the private entities were either publicly funded or established to carry out public business relating to the executive branch.[135]

GRAMA, however, did not compel disclosure from the Utah State Bar with regard to its arguably quasi-public activities, because the language of the statute only encompassed specific agencies of the judicial branch. GRAMA coverage was limited to "courts, the Judicial Council, the Office of the Court Administrator, and similar administrative units in the judicial branch."[136]

The 1992 edition of GRAMA, however, defines governmental entity more broadly. Rather than limiting broad disclosure to entities associated with the executive branch, GRAMA now compels disclosure from "every office, agency, board, bureau, committee, department, advisory board, or commission of the entities listed in subsection 9(a) [which includes courts] that is funded or established by the government to carry out the public's business."[137] This revision arguably extends GRAMA's disclosure requirements to the Utah State Bar, which has been officially recognized as "rendering advisory services" to the Utah Supreme Court.[138]

GRAMA requires disclosure of an employee's "gross compensation."[139] Since GRAMA may now apply to the Utah State Bar, one might argue that bar employee salary dates should now be open to the public. This result, however, would be unjust because bar employees do not receive public funds. In order to provide accountability for the exercise of public power, yet insure privacy where employees receive private funds, bar employees should not

---

134. *Id.* § 63-2-103(6)(a)(i).

135. Note that GRAMA does not fix a minimum percentage on the amount of funding needed to subject an entity to disclosure, as do Arkansas and Michigan statutes. *See supra* note 104.

136. UTAH CODE ANN. § 63-2-103(6)(a)(iii) (Supp. 1992).

137. UTAH CODE ANN. § 63-2-103 (Supp. 1992).

138. *Barnard*, 804 P.2d at 529-30.

139. UTAH CODE ANN. § 63-2-301 (Supp. 1992).

be compelled to disclose their salaries to the public. Instead, bar employees should be entitled to a protective order, as are judicial records.[140]

V. THE SEPARATION OF POWERS DOCTRINE

The doctrine of separation of powers may prevent legislatures from applying open records laws to state bars.

> The separation of powers doctrine arises from the concept that each branch of government is suited to make certain types of decisions; therefore, each branch should not intrude into the decision-making power of the other branches of government. The doctrine further attempts to limit the power of the branches of government so that one branch will not become too powerful and usurp power from a co-equal branch of government.[141]

Some state constitutions provide that the state supreme court has the complete authority to regulate the practice of law. Legislatures in such states are not able to regulate the practice of law without the supreme court's consent.

A. Washington, Minnesota, & Oregon

In *Graham v. Washington State Bar Association,*[142] the Washington Supreme Court held that even if the legislature intended to allow the State Auditor to audit the state bar, the doctrine of separation of powers prevented the state legislature from interfering with the court's power to regulate the bar.[143] The regulation of the practice of law, the court reasoned, was within the "sole jurisdiction" of the state supreme court.[144] Moreover, the court noted that the legislature's characterization of

---

140. UTAH CODE ANN. § 63-2-301(1)(f) (Supp. 1992); *supra* note 14 (authorities discussing protective orders in the context of judicial and bar disciplinary records).

141. Kent R. Hart, Note, *Court Rulemaking in Utah Following the 1985 Revision of the Utah Constitution*, 1992 UTAH L. REV. 153, 175 (citations omitted).

142. 548 P.2d 310 (Wash. 1976).

143. *Id.* at 315-16.

144. *Id.* at 315 (quoting State *ex rel.* Schwab v. Washington State Bar Ass'n, 493 P.2d 1237, 1238-39 (Wash. 1972)).

the bar as a state agency did not deprive the court of its right to control the bar.[145]

Similarly, in *Sharood v. Hatfield*,[146] the Minnesota Supreme Court struck down as unconstitutional a legislative attempt to channel attorney registration fees into the state's general revenue fund.[147] The court stated that "when the legislature attempts to go beyond merely indicating what it deems to be desirable, we have not hesitated to strike down such acts as unconstitutional."[148]

In contrast, in *Sadler v. Oregon State Bar*,[149] the Oregon Supreme Court upheld a state disclosure law which conflicted with a court rule prohibiting disclosure.[150] The court reasoned that: (1) governmental functions necessarily overlap;[151] (2) the overlap "arises from . . . legislative police power";[152] (3) disclosure did not "unreasonably encroach upon the judicial function of disciplining lawyers";[153] and (4) the legislature had already passed many laws regulating the state bar.[154]

Sadler can be distinguished from *Graham* and *Sharood*, however, because Oregon's constitution specifically allows the legislature to enact legislation affecting the judicial branch.[155] The Oregon Constitution provides that "[t]he courts, jurisdiction, and judicial system of Oregon, except so far as expressly changed by this amendment, shall remain as at present constituted *until otherwise provided by law.*"[156] Therefore, whether the doctrine of separation of powers prevents the legislature from applying state open records laws to state bars depends on whether a particular state constitution allows the legislature to regulate the judicial branch.

---

145. *Id.* at 316.
146. 210 N.W.2d 275 (Minn. 1973)
147. *Id.* at 282.
148. *Id.* at 279.
149. 550 P.2d 1218 (Or. 1976).
150. *Id.* at 1221, 1227.
151. *Id.* at 1222.
152. *Id.* at 1222-23.
153. *Id.* at 1226-27.
154. *Id.* at 1223-26.
155. *See id.* at 1223.
156. OR. CONST. art. VII, § 2 (emphasis added).

B. Utah

The *Barnard* court noted that the case raised the issue of whether Utah's disclosure acts violated the principle of separation of powers by regulating the judiciary.[157] However, the court specifically refrained from deciding whether the statutes would violate separation of powers if they were applied to the Utah State Bar.[158] Utah's constitution grants authority to "govern the practice of law" exclusively to the court.[159] In 1985, the Utah Legislature expanded the court's powers in this regard by amending the Utah Constitution.[160] Prior to the amendment, the court regulated the practice of law through its inherent authority, a fact recognized by the legislature in a 1931 statute.[161] The 1985 amendment added the italicized language to the following provision: "[t]he supreme court by rule shall govern the practice of law, *including admission to practice law and the conduct and discipline of persons admitted to practice law.*"[162]

Moreover, the Utah Supreme Court has said that "[a]lthough the legislature has some power to regulate and control attorneys, at least in certain respects, that power is subject to this Court's inherent power to discipline its officers."[163] Therefore, given the history of the legislature's recognition of the court's powers, and given the court's emphasis that it has the inherent power to regulate the practice of law, the Utah Supreme Court would probably hold that legislative attempts to regulate the practice of law violate the separation of powers doctrine.

In addition, if the legislature intended Utah's open records act to apply to the Utah State Bar, it could have included the bar as

---

157. Barnard v. Utah State Bar, 804 P.2d 526, 527 (Utah 1991).

158. *Id.* at 530.

159. *See* UTAH CONST. art. VIII, § 4.

160. *Barnard*, 804 P.2d at 528; S.J.R. Res. 1, 45th Leg., 2d Spec. Sess., 1984 Utah Laws 269.

161. *See Barnard*, 804 P.2d at 528.

162. UTAH CONST. art. VIII, § 4 (emphasis added). The purpose for amending the constitution is not clear if the court had inherent power to regulate the practice of law "from the beginning." *See Barnard*, 804 P.2d at 528. Inherent in the judicial power conferred on the Utah Supreme Court by article VIII, § 4 of the Utah Constitution is the "power to regulate the 'practice of law.'" *Id.* The amendment represents express legislative recognition that the supreme court, not the legislature, regulates the practice of law.

163. *In re* Disciplinary Action of McCune, 717 P.2d 701, 705 (Utah 1986).

a state agency under the statute, or otherwise specifically stated that Utah's open records laws apply to the bar.[164] However, as was evident from *Graham v. Washington Bar Ass'n,* there is no guarantee that such legislation would be upheld by the court.[165]

It is also important to consider whether applying disclosure laws to the Utah State Bar interferes with the Utah Supreme Court's power to regulate the practice of law. The court's current policy is to keep disciplinary proceedings confidential unless and until formal charges are filed.[166] Obviously, any attempt to apply open records legislation to the Utah State Bar's disciplinary activities, before formal charges are filed, would conflict with the court's rules.[167] Therefore, action by the Utah Legislature to apply open records laws to informal charges would violate the separation of powers doctrine and would not be enforceable without the Utah Supreme Court's consent.[168]

VI. DISCLOSURE OF BAR DISCIPLINARY RECORDS

Twenty-six jurisdictions, including Utah,[169] allow public access to bar disciplinary proceedings once formal charges of attorney misconduct are filed.[170] An issue currently subject to

---

164. For example, a North Carolina statute provides: "[t]here is hereby created as an agency of the State of North Carolina, for the purposes and with the powers hereinafter set forth, the North Carolina State Bar." N.C. GEN. STAT. § 84-15 (1985). In response to this legislation the North Carolina Supreme Court held that the bar was a state agency. *See* North Carolina State Bar v. Hall, 238 S.E.2d 521, 521 (N.C. 1977).

165. *See* Graham v. Washington State Bar Ass'n, 548 P.2d 310, 316 (Wash. 1976) (holding legislature's characterization of state bar as agency of state was unwarranted and an unconstitutional interference with power of judiciary).

166. *See* PROCEDURES OF DISCIPLINE OF THE UTAH STATE BAR, Rules VI, VII, XII (1991).

167. For example, the plaintiff in *Sadler* admitted that Oregon's open records statute when applied to the bar conflicted with the court's rules. *See* Sadler v. Oregon, 550 P.2d 1218, 1221 (Or. 1976).

168. One might argue that GRAMA does not pose a separation of powers problem regarding the bar's disciplinary records. GRAMA allows governmental entities to classify disciplinary and licensing records as protected if certain criteria re met. UTAH CODE ANN. § 63-2-304(8) (Supp. 1992). Criteria which might be cited in an argument against disclosure include: release of records interferes with investigations or disciplinary proceedings; records "certain a personal recommendation concerning an individual [and] disclosure would constitute a clearly unwarranted invasion of personal privacy"; and "disclosure is not in the public interest." *Id.* § 63-2-304(8)(a)–(b), (24).

169. *See* PROCEDURES OF DISCIPLINE OF THE UTAH STATE BAR, Rules VI, VII, XII (1991).

170. ABA REPORT, *supra* note 6, at 88.

debate is whether *all* state bar disciplinary proceedings, including dismissed complaints, should be subject to public access.[171]

Striking the correct balance between the public's interest in disclosure and an attorney's interest in privacy requires an analysis of the policy reasons underlying disclosure of disciplinary records. Some commentators suggest that making all bar disciplinary proceedings public will improve the legal profession's image and increase public trust in attorneys.[172] Further, some authorities suggest that all bar disciplinary proceedings should be open to protect the public and because attorneys, as officers of the court, are accountable to the public.[173] Another important reason for providing public access to all bar disciplinary proceedings is to provide public accountability where a government function has been delegated to a state bar.[174] The issue is how to achieve these goals while protecting the reputations of innocent lawyers—assuming the reputations of innocent lawyers need protection.

A. The McKay Commission and Increasing Public Trust

In 1991, the American Bar Association's Commission on Evaluation of Disciplinary Enforcement ("McKay Commission") called upon state bars to make all records of attorney disciplinary proceedings ("except the work product of disciplinary counsel") open to the public unless a protective order is obtained to prevent disclosure.[175] The McKay Commission's research showed that

---

171. In Utah, disciplinary records in other professions, including medicine, are not subject to public access until after formal charges are filed. Telephone Interview with David Robinson, Director, Utah Division of Professional and Occupational Licensing (Feb. 14, 1992). In West Virginia, however, unmeritorious complaints are open to public access upon dismissal. *See* Daily Gazette Co. v. West Va. Bd. of Medicine, 352 S.E.2d 66, 71 (W. Va. 1986) ("[I]f the Board finds that probable cause does not exist, the public has a right of access to the complaint or other document setting forth the charges, and the findings of fact and conclusions of law supporting the dismissal.").

172. *See* ABA REPORT, *supra* note 6, at iv, 23, 24.

173. *See* Daily Gazette Co., 326 S.E.2d at 710.

174. *See supra* notes 18-22 and accompanying text (discussing governmental responsibilities handled by private actors).

175. ABA REPORT, *supra* note 6, at 23. The McKay Commission also recommended major surgery to state bar disciplinary proceedings, suggesting that bar commissioners divest themselves of disciplinary responsibilities, that state supreme courts appoint attorneys to serve as disciplinary judges. *See id.* at 19-20. Therefore, the best way to generate public confidence in the judicial disciplinary process may be to sever the

disciplinary systems of state bars are fair, but that "a high level of public distrust" exists primarily as a result of secret disciplinary proceedings.[176]

The report of the McKay Commission was a significant departure from the American Bar Association's 1970 Clark Report,[177] which recommended that disciplinary proceedings be confidential until "hearings have been held and the charges sustained by the trial authority."[178] The Clark Report assumed: (1) that public access to unproven allegations would injure the innocent lawyer's reputation; (2) that subsequent acquittal would fail to restore the innocent attorney's reputation; and (3) that it was more important to defend innocent lawyers' reputations than to notify the public.[179]

In response to the Clark Report's desire to protect innocent lawyers' reputations by preventing disclosure of unmeritorious complaints, the McKay Commission pointed to states, including West Virginia, Florida, and Oregon, which have opened all records[180] of their disciplinary systems and have not noticed any harm to the reputations of innocent lawyers.[181] The McKay Commission stated that "we find in Oregon, Florida, and West Virginia ample experience to demonstrate that public proceedings

---

disciplinary functions from the private functions and allow public access to the disciplinary body's records.

176. *Id.* at iv. The McKay Commission noted the following:

> The commission is convinced that secrecy in discipline proceedings continues to be the greatest single source of public distrust of lawyer disciplinary systems. Because it engenders such distrust, secrecy does great harm to the reputation of the profession. . . . The public does not accept the profession's claims that lawyers' reputations are so fragile that they must be shielded from false complaints by special secret proceedings.

Id. at 23.

177. AMERICAN BAR ASS'N, PROBLEMS AND RECOMMENDATIONS IN DISCIPLINARY ENFORCEMENT: AMERICAN BAR ASS'N SPECIAL COMMITTEE ON EVALUATION OF DISCIPLINARY ENFORCEMENT COMMITTEE ON EVALUATION OF DISCIPLINARY ENFORCEMENT (1970) [hereinafter "CLARK REPORT"].

178. *Id.* at 138. *See generally* Douglas R. Haddock, Note, *The Legal Profession's Attempt to Discipline Its Members: A Critique of the Clark Report*, 1970 UTAH L. REV. 611 (providing critical analysis of Clark Report).

179. CLARK REPORT, *supra* note 177, at 138-40.

180. "Florida and West Virginia's disciplinary records are open to the public when a charge is filed or a complaint is dismissed. Oregon's records are public when a complaint is made. These open disciplinary systems have proven that lawyers are not harmed by them." *Id.* at iv.

181. *See id.* at 25-26.

or public records of dismissed complaints do no harm to innocent lawyers' reputations. On the contrary, secrecy does great harm to the reputation of the profession as a whole."[182]

After making dismissed complaints public, the West Virginia Bar initially experienced a period of public interest, but that interest soon waned and the press stopped examining the records.[183] The president of the Oregon State Bar, however, testified before the McKay Commission that "[w]ithout access to complaint information the public is suspicious that lawyers are protecting their own."[184]

State bars may be more reluctant to implement the McKay Commission's disclosure recommendations than they were to accept the Clark Report.[185] Because the McKay Commission did not identify any actual misconduct or problems with attorney discipline systems, and only pointed to public frustration as the basis for its recommendation,[186] state bars may argue that implementing the McKay Commission's recommendation will not actually improve the profession. Also, the public may never be totally satisfied with the regulation or practice of law, no matter how much it is allowed to know about the process.

If public criticisms of the disciplinary system are legitimate, it is possible that better self-regulation, rather than more openness, would more successfully accomplish the goal of improving the legal profession's image. As Judge Cardozo stated, "[i]f the house is to be cleaned, it is for those who occupy and govern it, rather than for strangers, to do the noisome work."[187]

Finally, it is also possible that there is an anxiety toward the legal profession that will attach itself to whatever activity members of the public hear about. Allowing more access to disciplinary proceedings may not resolve this anxiety. Instead, it

---

182. ABA REPORT, *supra* note 6, at 25-26.

183. *Id.* at 24.

184. *Id.*

185. In fact, on February 4, 1992, the ABA House of Delegates tacitly rejected the McKay Commission's recommendation that meritless complaints be made public, adopting an amended version of the report which requires disclosure only when a disciplinary panel finds probable cause. *See* LEGISLATIVE HISTORY OF REPORT 119, 1992 MIDYEAR MEETING OF THE HOUSE OF DELEGATES OF THE AMERICAN BAR ASS'N, RECOMMENDATION 7 (Feb. 4, 1992).

186. ABA REPORT, *supra* note 6, at iv.

187. People *ex rel.* Karlin v. Culkin, 162 N.E. 487, 493 (N.Y. 1928).

may fuel witch hunts; for example, lawyer referral services generating lists of (possibly unmeritorious) complaints. Because the McKay Commission's report and a similar report by the Utah State Bar[188] do not identify any actual abuse of the disciplinary system,[189] opening all stages of disciplinary proceedings to the public may be nothing more than a quick-fix attempt to solve an age-old[190] anxiety that will not actually improve the discipline process.

There are no simple methods to ease the public's frustration with the legal profession, and disclosing all of the bar's records will not suddenly cause people to congratulate attorneys on their openness. Therefore, opening up all disciplinary proceedings after a complaint has been dismissed, without regard to the privacy interests, may not be worth the cost to innocent lawyers' reputations.[191] This may be so even though the McKay Commission maintained that "little or no significant harm will come to lawyers from the public disclosure of mere allegations."[192]

B. Daily Gazette *and Protection of the Public*

Another reason to allow public access to disciplinary proceedings is protection of the public. In *Daily Gazette Co. v. Committee on Legal Ethics*,[193] the Supreme Court of Appeals of West Virginia held that there is a constitutional right of access to attorney

---

188. *See* UTAH BAR REPORT *supra* note 6.

189. *See* ABA REPORT, *supra* note 6, at iv.

190. There is no easy way to satisfy the public with the image of the legal profession. However, this is not because people have not tried to find easy answers. Long before state bars were created to assist courts in regulating the practice of law, Shakespeare's infamous insurrectionist "Dick the Butcher" suggested the famous anecdote that "[t]he first thing we do, let's kill all the lawyers." WILLIAM SHAKESPEARE, THE SECOND PART OF KING HENRY THE SIXTH act 4, sc. 2. That the anecdote is understood today in a sense completely opposite from the way it was originally intended further illustrates the acuteness of modern society's general frustration with the legal profession. *See* Walters v. National Ass'n of Radiation Survivors, 473 U.S. 305, 371 n.24 (1985) (Stevens, J., dissenting).

191. *See* Michael Franck, *Public Disclosure of Innuendo, Unsubstantiated Allegations and Other Assaults on Reputations*, 70 MICH. B.J., Jan. 1991, at 12. ("No valid rationale exists for opening unsubstantiated allegations to public inspection.").

192. ABA REPORT, *supra* note 6, at 88. The McKay Commission found that no significant harm would come to the reputations of innocent lawyers "from the public disclosure of mere allegations." *Id.*

193. 326 S.E.2d 705 (W. Va. 1984).

disciplinary proceedings.[194] The court also invalidated a West Virginia State Bar bylaw preventing disclosure of disciplinary records unless public discipline was recommended by the bar's legal ethics committee.[195]

The *Daily Gazette* court rejected traditional notions that attorney discipline should be confidential until formal charges are filed[196] by holding that "once a complaint of unethical conduct in an attorney disciplinary proceeding is dismissed for lack of probable cause, the public has a right of access to the complaint and the findings . . . presented in support of such dismissal."[197] In support of its position, the court reasoned that "'the primary purpose of the ethics committee is not punishment but rather the protection of the public and the reassurance of the public as to the reliability and integrity of attorneys.'"[198] Furthermore, the court reasoned that "if the legal profession's practice of self-regulation is to remain viable, the public must be able to observe for [itself] that the process is impartial and effective."[199]

The *Daily Gazette* court also held that the public had an interest in attorney discipline because of the lawyer's role as an officer of the court.[200] As an officer of the court, the attorney is obligated to assist in the administration of law.[201] Also, the public's interest in disclosure is high when an attorney's misconduct results in formal charges being brought against the attor-

---

194. *Id.* at 711-12. The court also reasoned that the use of private reprimands as a method of official discipline contravenes the state's open courts provision because disciplining attorneys is for the benefit of the public, is the business of the public, and therefore should be carried on in public. *Id.* at 713-14; *see also In re* Matter of Johnson, 461 N.W.2d 767, 770 (S.D. 1990) (interest of public and media in access to attorney affidavit in attorney disciplinary proceeding outweighed interests in privacy).

195. *See Daily Gazette*, 326 S.E.2d at 709-11. Under the bylaw struck down in the case, the ultimate disposition of charges against an attorney "are generally never made a matter of public record, but are forever cloaked in a veil of secrecy." *Id.* at 709.

196. *Id.* at 712. Traditional notions are embodied by the 1970 Clark Report, which recommended that disciplinary proceedings be confidential until "hearings have been held and the charges sustained by the trial authority." CLARK REPORT, *supra* note 177, at 138.

197. *Daily Gazette*, 326 S.E.2d at 713.

198. *Id.* at 709 (quoting Committee on Legal Ethics v. Mullins, 226 S.E.2d 427, 428-29 (W. Va. 1976)).

199. *Id.* at 711.

200. *Id.* at 710. "[T]he principal purpose of attorney disciplinary proceedings is to safeguard the public's interest in the administration of justice." *Id.*

201. *Id.*

ney.[202] Therefore, when formal charges are brought, the public has a right to receive all of the disciplinary records of an officer of the court.[203]

The American Bar Association ("ABA") is concerned with the protection of the public once formal charges have been filed against an attorney.[204] The ABA Model Rules for Lawyer Disciplinary Enforcement provide:

> Once a finding of probable cause has been made, there is no longer a danger that the allegations against the respondent are frivolous. The need to assure the integrity of the disciplinary process in the eyes of the public requires that at this point further proceedings be open to the public. An announcement that a lawyer accused of serious misconduct has been exonerated after a hearing behind closed doors will be suspect. The same disposition will command respect if the public has had access to the evidence.[205]

Thus, both courts and commentators agree that once formal charges are filed, the public has an interest in disclosure of disciplinary records. The *Daily Gazette* court expanded on this concept, holding that even unmeritorious complaints should be made public.[206] The McKay Commission adopted this policy and recommended that all state bars follow it.[207]

On the other hand, the public's interest in disclosure is arguably not as high when a complaint is lodged but no formal charges result. Since the vast majority of the claims made against attorneys are frivolous,[208] the potential for injuring an attorney's reputation weighs heavily against allowing public access to all

---

202. *Id.* at 712.

203. *Id.* at 713.

204. ABA REPORT, *supra* note 6, at 87.

205. ABA MODEL RULES FOR LAWYER DISCIPLINARY ENFORCEMENT Rule 16 cmt. (1965).

206. *See Daily Gazette*, 326 S.E.2d at 713.

207. *See* ABA REPORT, *supra* note 6, at xii, 23, 88. The McKay Committee adopted this policy in its official 1991 report, but the recommendation was withdrawn before the American Bar Association's House of Delegates voted on the issue in February, 1992. *See* LEGISLATIVE HISTORY OF REPORT 119, *supra* note 185.

208. *See Daily Gazette*, 326 S.E.2d at 713 n.15 (quoting *The Bar and Watergate: Conversation with Chesterfield Smith*, 1 HASTINGS CONST. L.Q. 31, 35 (1974)); *see also infra* note 226 (640 investigations resulted in only 67 instances of discipline).

allegations. Therefore, despite the McKay Commission's confidence that innocent attorneys will not be hurt by publishing meritless complaints, state bars should consider ways to protect the reputations of attorneys while also disclosing as many records about the proceedings as possible.

C. Accountability for Exercising Delegated Powers

A third reason articulated for applying open records statutes to state bar disciplinary activities is that the powers to discipline are delegated from the government.[209] Although it is difficult to define "governmental function" or "governmental powers,"[210] bars carry out governmentally-delegated functions, including discipline of attorneys and administering the bar admissions process.[211] For example, the Utah State Bar has no actual authority to order disbarment, suspension, or to give a public reprimand.[212] However, it does have the authority to dismiss a complaint as unmeritorious[213] and to make recommendations to the Utah Supreme Court.[214]

State bar accountability to the public for exercising disciplinary powers granted by government is fundamentally different from the two other policy reasons for disclosure—increasing public trust in the legal profession[215] and protecting the public.[216] The state bar, not the disciplined attorney, is accountable to the public for its exercise of governmental power. The *Daily Gazette* court held that "once a complaint of unethical conduct . . . is dismissed for lack of probable cause, the public has a right of access to the complaint and the findings of fact and conclusions of law which are presented in support of such dismissal."[217] Thus, state bar

---

209. *See supra* notes 18-22 and accompanying text (discussing governmental responsibilities handled by private actors).

210. David M. Lawrence, *Private Exercise of Governmental Power*, 61 IND. L.J. 647, 647 n.1, 648 (1986) (defining "governmental function" characterized as "definitional thicket").

211. *Id.* at 648 (licensing is governmental function); Barnard v. Utah State Bar, 804 P.2d 526, 529 (Utah 1991) (discipline is governmentally delegated function).

212. *Barnard*, 804 P.2d at 529.

213. PROCEDURES OF DISCIPLINE OF THE UTAH STATE BAR, Rule 5(b)(2) (1991).

214. *Barnard*, 804 P.2d at 529.

215. *See supra* notes 175-192 and accompanying text.

216. *See supra* notes 193-208 and accompanying text.

217. *Daily Gazette*, 326 S.E.2d at 713.

disciplinary committees should account to the public on their exercise of governmental power by allowing access to their decisions.

However, there is a difference between applying open records laws to state bar disciplinary proceedings and applying open records laws to other delegations of governmental power, such as government hiring of an engineering firm or a towing agency.[218] Although an entity entrusted with public power should be accountable to the public for such power,[219] a high interest in privacy attaches to state bar disciplinary proceedings, while a lower interest in privacy applies to the records of a towing agency which has towed away automobiles. This is because disciplinary proceedings may deal with highly sensitive and personal issues involving matters not connected with an attorney's professional life.[220]

The key similarity, however, between the towing agency and the bar is that the state has delegated power to both entities, and both should be accountable to the public for the use of that power. If the state delegates power to a towing agency to tow automobiles and the state pays the towing agency public money, the towing agency should be accountable for the way it carries out its activities, just as any public agency would be accountable. Similarly, when the state delegates power to the bar to discipline attorneys, the bar should be accountable to the public for the way it uses its public power.

D. Preserving Privacy Yet Providing for Accountability

One way to protect the privacy interests of attorneys, while still providing for public accountability in carrying out disciplinary proceedings and generating public trust, is to compel disclosure only to the extent necessary to allow public accountability for the power delegated. For example, courts and legislatures could

---

218. *See supra* notes 20, 44-45 and accompanying text (discussing private actions as delegated functions).

219. *See* Westchester Rockland Newspapers, Inc. v. Kimball, 430 N.Y.S.2d 574, 576 (1980).

220. *See, e.g.*, Disciplinary Bd. v. Bergan, 592 P.2d 814, 818 (Haw. 1979) (violation of narcotics laws grounds for disbarment).

mandate public access to bar disciplinary records generally, but allow redaction of portions of the record identifying specific attorneys or otherwise intruding unduly into their private lives.

Thus, the media, lawyer referral services, and others interested in complaints against attorneys, would not be allowed to harm the reputation of innocent attorneys by asking for all the complaints dismissed by the disciplinary committee and then publishing the names and the allegations. Instead, the public and the media would have access to the disciplinary committee's reasons for dismissing complaints, but would not be able to identify the innocent attorney.

The Utah Supreme Court accepted a similar approach in *KUTV, Inc. v. Utah State Board of Education*,[221] in which a local Utah television station requested access to confidential questionnaires surveying religious discrimination in a Utah high school.[222] The court ruled that the confidential student questionnaires should be made public, unless the district court found that it would be "impossible to edit the questionnaire responses to preserve confidentiality."[223] The court held that although keeping public records open for inspection prevents secrecy in public affairs, the public's right of access is limited by an "implied rule of reason."[224] The court then held that the trial court should view the documents *in camera* and determine whether it is possible to edit the responses to preserve the confidentiality which the students had been promised before answering the questionnaire.[225]

Courts and legislatures should be subject to the rule of reason regarding access to attorney disciplinary proceedings and should balance the public's interest in disclosure against an attorney's interest in privacy. Redacting the attorneys' names and sensitive personal information before disclosing unmeritorious complaints satisfactorily balances the interests of disclosure and privacy because: (1) state bars do not receive public funds with which to

---

221. 689 P.2d 1357 (Utah 1984).

222. *Id.* at 1358.

223. *Id.* at 1362.

224. *Id.* at 1361-62 (citing Bruce v. Gregory, 423 P.2d 193, 199 (Cal. 1967)). The "implied rule of reason" weighs the presumption of access against an individual's interest in confidentiality.

225. *Id.* at 1362.

conduct disciplinary proceedings; (2) the purpose of disclosure is to generate confidence in the disciplinary process, not create a public inquisition against individual attorneys; (3) the McKay Commission and the Utah State Bar have found no foul play in the current disciplinary process;[226] and (4) attorneys, though they are officers of the court, have more of an expectation of privacy than paid public officers.

One might argue that redacting identifying information is impractical and inefficient.[227] Should the bar pay for redaction, or should the individual seeking disclosure be required to pay? Second, should the state bar, which must defend lawsuits with bar member funds, be open to liability for mistakes in redaction? Third, there is the practical consideration of whether volunteer bar commissioners[228] should be subject to additional scrutiny and potential lawsuits resulting from increased disclosure. Finally, redaction of too many facts could make the disclosure meaningless. Thus, although redaction might prove useful in some cases, it may be impractical as a general policy.

Regarding disclosure of unmeritorious complaints, state bars have many options, none of which completely satisfies the interests of public disclosure, privacy, and efficiency in every situation: (1) follow the McKay Commission's recommendation and disclose all records, subject to an attorney's right to obtain a protective order; (2) follow the Clark Report and disclose no

---

226. ABA REPORT, *supra* note 6, at iv; UTAH BAR REPORT, *supra* note 6, at 19. In fact, the Utah State Bar's disciplinary system is arguably more strict than the state's disciplinary practices for other professionals:

> The present Bar discipline process was found to be well-funded and vigorous. In 1987, direct costs of discipline were $118,458. . . .
>
> In 1989, the Bar received approximately 800 complaints about lawyers . . . . Of the 640 investigations conducted in 1990, there were 25 instances of informal discipline, and 27 instances of public discipline, including 13 suspensions and 2 disbarments, for a suspension/disbarment ratio of 2.94 per 1000 licensees. By comparison, the Division of Occupational and Professional Licensing investigated 1,094 complaints on its 85,000 licensees (including such professions as doctors and accountants), handing out 12 suspensions and 11 revocations . . . for a suspension/revocation ratio of 0.27 per 1000 licensees.

UTAH BAR REPORT, *supra* note 6, at 19 (citations omitted).

227. *See supra* notes 221-26 and accompanying text (discussing public's right to access limited by implied rule of reason).

228. "The Bar is governed by an eleven member Board of Commissioners" and, "[t]he Bar has a full-time Executive Director, with a paid staff of 15 other persons. Commissioners and elected officials serve without pay." UTAH BAR REPORT, *supra* note 6, at 8-9.

unmeritorious complaints; or (3) adopt a combination of the Utah Supreme Court's balancing test in *Redding* and the court's protection of privacy in *KUTV,* disclosing the records but first redacting identifying information. Adopting the McKay Commission's original disclosure recommendations would provide an effective check on governmentally-delegated powers, while still allowing courts to issue protective orders against disclosure when privacy interests should prevail. In determining whether to issue a protective order, courts should perform a *Redding*-type balancing to determine when privacy interests outweigh the interests of public disclosure. A final option strikes a reasonable balance between the public's interest in accountability for delegated power and an individual attorney's interest in privacy. The Utah Bar Journal provides some information on bar disciplinary activities, including information on admonitions, suspensions, and disbarments.[229] The information given sometimes excludes the names of the individuals involved, but still satisfies the public's interest in access because it allows the public to know why the state bar decides to take certain disciplinary measures and decides against others.[230]

Taking a similar approach to dismissed complaints may be a satisfactory compromise between the competing interests in this area. Making such information available to the public would allow the public to know the reasons behind the bar's decision, yet would protect attorney's privacy interests as well.

E. Public Access to Bar Admission Proceedings

Another currently debated issue is whether the public has a right of access to bar admission proceedings. For example, in *Pasik v. State Board of Law Examiners,*[231] the court held that the board of law examiners exercised a judicial function and therefore

229. *See, e.g., Discipline Corner,* UTAH B.J., Aug./Sept. 1991, at 22.

230. *See id.* For example, in one edition, the Utah Bar Journal published the following under the title "Admonitions": "2. An attorney was admonished for violating Rule 3.5a (Decorum). The attorney recorded a telephone conversation between himself and a judge without the judge's knowledge or consent. The discipline was mitigated in that the attorney acknowledged the misconduct and apologized to the judge prior to the issuance of discipline." *Id.*

231. 478 N.Y.S.2d 270 (App. Div. 1984).

was a "court" within the meaning of an exemption allowed to courts from the state's freedom of information act.[232] Thus, the board had no duty to disclose information connected with the bar examination.[233]

However, in *Connecticut Bar Examining Committee v. Freedom of Information Commission*,[234] the Connecticut Supreme Court held that although the bar examining committee is an arm of the court, it carries out some administrative functions which should be subject to public access.[235] For example, promulgating the criteria for admission and publishing the rules of practice are administrative rather than judicial functions.[236] The court held that records pertaining to administrative functions should be subject to disclosure unless disclosure would encroach upon a judicial function of the committee.[237]

Since the admissions process is a governmentally-delegated function, the admissions committee should be subject to the checks and balances of public disclosure; but only to the extent necessary, because admissions committees may discuss information which is confidential. However, bar admissions committees carry out some administrative functions in which attorneys and members of the public have an interest. Disclosing matters of an administrative nature would arguably not infringe upon privacy concerns. For example, the public may have an interest in general information regarding administrative procedures, and how exam questions are formulated and graded. There is no privacy interest to be protected in revealing these materials. Finally, although privacy concerns involved in an applicant's moral fitness to sit for the bar exam may require redaction of identifying information, the public may have an interest in knowing why attorneys are admitted or not admitted to the bar because of character issues.

---

232. *Id.* at 273. The case involved a conscientious young attorney who had recently passed the bar and requested information regarding his score on each section of his examination, and his percentile ranking on the examination. *See id.* at 270-71.

233. *Id.* at 273-74.

234. 550 A.2d 633 (Conn. 1988).

235. *Id.* at 636.

236. *See id.*

237. *Id.*

VII. STATUTES COMPELLING DISCLOSURE FROM STATE BARS

Two final issues remain. First, whether it should be courts or legislatures who compel disclosure; and second, if left to the legislatures, how the legislatures should go about the task.

One advantage to having courts apply their own rules and compelling access from state bars is that the bars can avoid the attorney fee and penalty provisions of state disclosure statutes.[238] Furthermore, courts can regulate the practice of law by regulating disclosure. Courts can prevent state legislatures from applying disclosure statutes to state bars because state supreme courts perform a dual role in determining whether a state disclosure law (or any other law) should apply to state bars. First, the court typically promulgates the rules for the bar—or delegates their promulgation to the bar subject to court approval.[239] Second, the court determines whether or not legislative enactments conflict with the rules and the separation of powers doctrine.[240] Thus, subject only to the scrutiny of the United States Supreme Court, and possibly a constitutional amendment, some state supreme courts have unchecked power to promulgate their own rules. Therefore, courts rather than legislatures should take the lead in compelling disclosure from state bars.

However, if courts do not make the first move, state legislatures can at least evince a legislative opinion that state bars should disclose their records. Legislatures, in attempting to apply disclosure laws to state bars, should consider several issues. *Keller* held that the California State Bar is a private agency,[241] and the *Graham* court held that the Washington State Bar is a private agency, despite the fact that the Washington legislature designated it as public.[242] Thus, legislative goals may be better served by applying disclosure acts to private agencies, rather than attempting to designate state bars as public agencies.

However, if legislatures choose not to pursue this option, other options are available. Legislatures could employ a broad,

---

238. *See, e.g.* ARK. CODE ANN. § 25-19-104, -107 (1992).

239. *See* Barnard v. Utah State Bar, 804 P.2d 526, 529 (Utah 1991).

240. *See supra* notes 141-68 and accompanying text (discussing unconstitutionality of attempts to legislate in this area).

241. *See* Keller v. State Bar of California, 110 S.Ct. 2228, 2234-35 (1990).

242. Graham v. Washington State Bar Ass'n, 548 P.2d 310, 316 (West 1976).

flexible statute, such as Utah's GRAMA,[243] when applying state open records laws to the public functions carried out by state bars. Under this type of statute, private entities are subject to disclosure if they carry out delegated public functions or receive public funds.[244] One problem with such a statute is that it does not protect the private aspects of state bars. Florida's statute offers a possible solution.[245] The Florida statute protects privacy interests but would still apply to the bar insofar as the bar carries out governmental functions.[246] The problem with Florida's statute is that perhaps it is too broad and flexible and, is likely to create more litigation over which functions are private and which are public.[247]

Possibly the best method of applying disclosure laws to private agencies which carry out governmental functions is to apply the law only to the specific function carried out by the entity. Utah's open meetings statute regarding attorney discipline is a good example.[248] The statute provides that "[a]ny Bar proceeding to consider formal charges against any member of the Bar shall be subject to the provisions of Title 52, Chapter 4 regarding open and public meetings."[249]

Oregon has adopted a similar approach. Oregon's State Bar statute provides that

---

243. UTAH CODE ANN. §§ 63-2-101 to -909 (Supp. 1992).

244. *Id.* § 63-2-103.

245. *See supra* note 44 and accompanying text (discussing Florida's open records statute).

246. Fritz v. Norflor Constr. Co., 386 So. 2d 899, 901 (Fla. Dist. Ct. App. 1980) (performing services for city requires "agency" to disclose "public records").

247. The New York Court of Appeals rejected a volunteer fire department's attempt to separate the records of its nongovernmental functions from its governmental functions in order to prevent disclosure of its nongovernmental records. Westchester Rockland Newspapers, Inc. v. Kimball, 430 N.Y.S.2d 574, 577 (1980). The court reasoned that nothing in the state's statutory definition of "record" turned on the purpose for which the document was produced, and to draw a line between the department's private and public functions was too difficult. *Id.* at 577-78.

248. UTAH CODE ANN. § 78-51-12 (1992). Utah's open meeting laws only apply to the meetings of a "public body." *Id.* § 52-4-2 (1989). The fact that the open meetings law applies to Utah State Bar disciplinary proceedings is some evidence that, even before GRAMA, the legislature considered at least part of the bar's activities to be "public."

249. UTAH CODE ANN. § 78-51-12 (1992). This statute reflects a Clark Report policy of keeping disciplinary proceedings secret unless formal charges are filed. This is the current policy of the Utah Bar Association. PROCEDURES OF DISCIPLINE OF THE UTAH STATE BAR, Rules VI-VII (1991).

> the Oregon State Bar is a public corporation and an instrumentality of the Judicial Department of the government of the State of Oregon . . . . The Bar is subject to the following statutes applicable to public bodies. . . . However, the bar is not subject to any other statute applicable to a state agency, department, board or commission or public body unless such statute expressly provides that it is applicable to the Oregon State Bar.[250]

The advantage to this line-item method of specifically designating the circumstances in which a state bar is subject to an act is that the legislative intent is clear because the legislature designates the activities it intends the disclosure law to cover. There is a decreased chance for litigation as a result of broad legislation. The disadvantage, however, is that some governmental functions may fall through the cracks and remain untouched by disclosure requirements.

Oregon's and Utah's "line item" approaches may be the best way to provide for public access to the bar's governmentally-delegated activities while preventing access to the bar's legitimately private ones. However, these approaches do not guarantee that no litigation will result to clarify the activities to which the statute applies. Although one might expect a specific statute such as Oregon's to curtail litigation, attorneys invariably find a way to litigate some aspect of the statute. For example, the specific Oregon statute quoted above was the subject of litigation in *State ex rel. Frohnmayer v. Oregon State Bar.*[251] In *Frohnmayer*, the court clarified the statute by holding that the Oregon State Bar was a "state agency" for purposes of the disclosure statute, rather than a "public agency of the state."[252] The net result is that under Oregon law, the state attorney general, rather than the district attorney, interprets the state disclosure statute as it applies to the bar.[253] Nevertheless, no statute is litigation-proof, and the line-item method appears to be the best way for the legislature to express that open records laws should apply to publicly-delegated activities, but not to private activities.

---

250. OR. REV. STAT. § 9.010(1) (1991).
251. 756 P.2d 689 (Or. Ct. App. 1988).
252. *Id.* at 691.
253. *See id.*

VIII. CONCLUSION

Barnard carves out a significant exception to the cases liberally allowing public access to state bar records. One might argue that in light of *Daily Gazette*,[254] *Sadler*,[255] and Florida's recent move to an open disciplinary process,[256] the *Barnard* decision represents a step backward to the era of secrecy. *Barnard* is not a throwback to the former era, however, because it represents a situation in which there was little public interest in access, yet a significant interest in privacy.

Courts and legislatures should compel public access from state bars because they carry out governmentally-delegated functions, not merely because they hope to improve the legal profession's public image. Applying disclosure laws to the public functions of the bar, while still retaining respect for privacy, is a complex problem and may result in confusion and litigation, but it balances the interests of bar members, of the bar's employees, and of the public.

DAVID B. DELLENBACH

---

254. *See supra* notes 193-208 and accompanying text (discussing *Daily Gazette*, allowing disclosure once formal complaint filed or dismissed).

255. *See supra* notes 149-56 and accompanying text (discussing *Sadler*, upholding disclosure law which conflicted with court rule).

256. *See supra* notes 44-45 and accompanying text (discussing Florida disclosure statute).

Johnson v. Morton Thiokol and Handbook Disclaimers: Allowing Employers to Have Their Cake and Eat It Too

> A principle of law that produces justice in a finite factual realm produces injustice when overextended.[1]

I. INTRODUCTION

Utah recently joined the growing number of jurisdictions that recognize limitations on employers' absolute right to terminate employees.[2] This recognition signals a departure from the traditional rule that allowed an employer to terminate an employee for "'no cause, good cause, or even cause morally wrong without fear of liability.'"[3] In particular, the Utah Supreme Court recognized an implied contract exception to the employment-at-will rule.[4] According to the implied contract exception, an employer's promissory representations, including those found in employee handbooks,[5] may be used to limit the employer's absolute discretion to terminate an employment relationship.[6]

---

1. Walter E. Oberer, *On Law, Lawyering, and Law Professing:* The Golden Sand, 39 J. LEGAL EDUC. 203, 205 (1989). The author would like to thank Professor Oberer for his helpful comments and criticism in preparing this Note for publication. The opinions expressed, of course, are those of the author.

2. Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1044-45 (Utah 1989) (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J., concurring); *id.* at 1052-53 (Zimmerman, J., concurring).

3. Rose v. Allied Dev. Co., 719 P.2d 83, 84-85 (Utah 1986) (quoting Thompson v. St. Regis Paper Co., 685 P.2d 1081, 1085 (Wash. 1984)).

4. Howcroft v. Mountain States Tel. & Tel. Co., 712 F. Supp. 1514, 1519 (D. Utah 1989) (applying Utah law); Johnson v. Morton Thiokol, Inc., 818 P.2d 997, 1000 (Utah 1991); Brehany v. Nordstrom, Inc., 812 P.2d 49, 54 (Utah 1991); Arnold v. B.J. Titan Serv. Co., 783 P.2d 541, 543-44 (Utah 1989); Lowe v. Sorenson Research Co., 779 P.2d 668, 670 (Utah 1989); Caldwell v. Ford, Bacon & Davis Utah, Inc., 777 P.2d 483, 485 (Utah 1989); *Berube*, 771 P.2d at 1044-45 (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J., concurring); *id.* at 1052 (Zimmerman, J., concurring).

5. For clarity and convenience, this Note uses the term "handbook" to refer to any written statement of an employer's policy that is distributed to employees, including "employee manuals, bulletins, and the like." *Berube*, 771 P.2d at 1052 (Zimmerman, J., concurring).

6. *See Howcroft*, 712 F. Supp. at 1519; *Johnson*, 818 P.2d at 1000; *Brehany*, 812 P.2d at 56; *Arnold*, 783 P.2d at 543-44; *Lowe*, 779 P.2d at 670; *Caldwell*, 777 P.2d at 485-86; *Berube*, 771 P.2d at 1044-45 (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J.,

In an effort to defeat implied contracts and maintain at-will employment, employers increasingly have resorted to disclaiming contractual liability in their employee handbooks.[7] Courts considering these disclaimers are split on their enforceability.[8] The Utah Supreme Court recently considered the enforceability of handbook disclaimers in *Johnson v. Morton Thiokol, Inc.*,[9] upholding the disclaimer despite the rather detailed termination and grievance procedures that accompanied it.[10] Unfortunately, *Johnson*'s broad holding encourages employers to avoid wrongful termination litigation simply by incorporating disclaimers into their handbooks, rather than by promulgating and abiding by reasonable employment policies.[11]

This Note sets out an analytical framework for determining the enforceability of disclaimers in employee handbooks. Part II examines the historical background of employment-at-will and the development of judicially created exceptions to the doctrine. Based on this history, Part III outlines the current split of authority on the enforceability of handbook disclaimers. Part IV then critically examines the *Johnson* decision and proposes an alternative framework for balancing employers' interests in maintaining at-will employment, employees' interests in being able to rely on promises of job security, and the public's interest in avoiding wasteful wrongful termination litigation.

---

concurring); *id.* at 1052 (Zimmerman, J., concurring).

7. *See* Michael A. Chagares, *Utilization of the Disclaimer as an Effective Means to Define the Employment Relationship*, 17 HOFSTRA L. REV. 365, 376 (1989); Patricia M. Lenard, Note, *Unjust Dismissal of Employees at Will: Are Disclaimers A Final Solution?*, 15 FORDHAM URB. L.J. 533, 564 (1987).

8. *See infra* notes 174-224 and accompanying text (discussing split of authority on enforceability of disclaimers).

9. 818 P.2d 997, 1003 (Utah 1991).

10. *Id.* at 1003-04. Chief Justice Hall wrote the majority opinion in *Johnson*, in which Justices Howe and Zimmerman joined. *Id.* at 1004. Justice Stewart wrote a concurring opinion and was joined by Justice Durham. *Id.* at 1004-06 (Stewart & Durham, JJ., concurring).

11. *See* Charlotte L. Miller, *Recent Developments in Utah Employment Law*, UTAH B.J., Oct. 1991, at 7, 11 & n.5 (advising Utah employers to use handbooks as management rather than litigation tools).

II. BACKGROUND

A. Development of the At-Will Rule

Early American courts generally followed the English common law presumption that an indefinite hiring was for one year.[12] The combined effects of the industrial revolution and laissez-faire economics, however, led American courts to depart from the English rule in the late nineteenth century.[13] Instead, American courts began to follow the so-called American rule. The rule generally is attributed to H.G. Wood's treatise on master and servant,[14] where he stated:

> With us the rule is inflexible, that a general or indefinite hiring is *prima facie* a hiring at will, and if the servant seeks to make it out a yearly hiring, the burden is upon him to establish it by proof. A hiring at so much a day, week, month or year, no time being specified, is an indefinite hiring, and no presumption attaches that it was for a day even, but only at the rate fixed for whatever time the party may serve.[15]

Recently, commentators have attacked Wood's statement of the rule, claiming it is unsupported by the cases he cited.[16] Neverthe-

---

12. *See, e.g.*, Sanford M. Jacoby, *The Duration of Indefinite Employment Contracts in the United States and England: An Historical Analysis*, 5 COMP. LAB. L. 85-128 (1982), *reprinted in* WILLIAM J. HOLLOWAY & MICHAEL J. LEACH, EMPLOYMENT TERMINATION RIGHTS AND REMEDIES 3, 8 (1985) (in-depth historical discussion of English and American development of employment-at-will); J. Peter Shapiro & James F. Tune, Note, *Implied Contract Rights to Job Security*, 26 STAN. L. REV. 335, 340-41 (1974).

13. *See* Lenard, *supra* note 7, at 538-39; Shapiro & Tune, *supra* note 12, at 340-41.

14. *See, e.g.*, Jacoby, *supra* note 12, at 27; Cornelius J. Peck, *Penetrating Doctrinal Camouflage: Understanding the Development of the Law of Wrongful Discharge*, 66 WASH. L. REV. 719, 722 (1991); Shapiro & Tune, *supra* note 12, at 341.

15. HORACE G. WOOD, A TREATISE ON THE LAW OF MASTER AND SERVANT § 136, at 283 (2d ed. 1886).

16. *See* Shapiro & Tune, *supra* note 12, at 341-43. In Wood's defense, Professor Jacoby points out that Wood's rule is somewhat misunderstood. Modern commentators argue that Wood's rule established absolute at-will employment. Jacoby explains, however, that although Wood's rule was generally applicable, it did not apply if the employee could demonstrate that the parties intended the employment to continue for a specific period or until the occurrence of a certain event. Jacoby, *supra* note 12, at 27. Thus, according to Wood, a court could look at all of the circumstances surrounding the employment relationship to determine the duration and conditions of employment. *See id.*; *see also* Matthew W. Finkin, *The Bureaucratization of Work: Employer Policies and Contract Law*, 1986 WIS. L. REV. 733, 735 (Wood's rule was presumption that could be

less, the rule was well suited for nascent industry, and nineteenth-century courts willingly adopted it in lieu of the one-year hiring presumption.[17]

The at-will rule assumed constitutional dimensions in *Adair v. United States*.[18] In *Adair*, the Supreme Court struck down a statute prohibiting employers from discharging employees because of union membership.[19] The Court stated, "the right of the employ[ee] to quit the service of the employer, for whatever reason, is the same as the right of the employer, for whatever reason, to dispense with the services of such employ[ee]."[20] Thus, by the early part of the twentieth century, the employment-at-will doctrine was firmly entrenched in American jurisprudence.[21]

B. Development of Exceptions to the At-Will Rule

Application of the at-will rule, however, often led to harsh results and, as a consequence, courts began to develop exceptions to the rule. The first break in the at-will rule came in *Petermann v. International Brotherhood of Teamsters Local 396*.[22] In *Petermann*, the Teamsters had hired Petermann as a business agent.[23] Petermann claimed that the Teamsters demanded he give false statements to a state legislative committee;[24] the day after Petermann testified truthfully, the Teamsters terminated him. The California District Court of Appeals noted that because Peter-

---

rebutted by evidence of trade practices and other circumstances of each case).

17. *See* Peck, *supra* note 14, at 722.

18. 208 U.S. 161 (1908).

19. *Id.* at 174-75.

20. *Id.*; *see also* Coppage v. Kansas, 236 U.S. 1, 26 (1915) (Court invalidated state statute outlawing discharge for union membership). Although the Court never expressly overruled *Adair* and *Coppage*, later cases indicated that they were no longer controlling. *See, e.g.*, Phelps Dodge Corp. v. NLRB, 313 U.S. 177, 187 (1941) ("The course of decisions in this Court since [*Adair* and *Coppage*] have completely sapped those cases of their authority.").

21. The at-will rule became so firmly rooted that an *American Law Reports* annotation concluded that "few legal principles would seem to be better settled than the broad generality that an employment for an indefinite term is regarded as an employment at will which may be terminated at any time by either party for any reason or for no reason at all." W.E. Shipley, Annotation, *Employee's Arbitrary Dismissal as Breach of Employment Contract Terminable at Will*, 62 A.L.R.3D 271, 271 (1975).

22. 344 P.2d 25 (Cal. Dist. Ct. App. 1959).

23. *Id.* at 26.

24. *Id.*

mann's employment contract was not for a specified duration, he was an at-will employee. Nonetheless, the court found that "the right to discharge an employee under such a contract may be limited . . . by considerations of public policy."[25] The court continued, "[i]t would be obnoxious to the interests of the state and contrary to public policy and sound morality to allow an employer to discharge any employee . . . on the ground that the employee declined to commit perjury."[26]

Although *Petermann* initially was considered an anomaly,[27] over time courts began to recognize wrongful discharge claims based on three general exceptions to the at-will rule: (1) the implied covenant of good faith and fair dealing; (2) public policy; and (3) implied-in-fact contracts. Each of these exceptions is briefly summarized below.

1. Implied Covenant of Good Faith and Fair Dealing

The implied-in-law covenant of good faith and fair dealing[28] prohibits either party to an agreement from interfering with the other party's right to receive the benefit of the agreement.[29] Most

---

25. *Id.* at 27.

26. *Id.*

27. *See* Peck, *supra* note 14, at 723-24.

28. For an in-depth discussion of the implied covenant of good faith and fair dealing exception to the employment-at-will rule, see Note, *Protecting At Will Employees Against Wrongful Discharge: The Duty to Terminate Only in Good Faith*, 93 HARV. L. REV. 1816 (1980).

29. Fortune v. National Cash Register Co., 364 N.E.2d 1251, 1257 (Mass. 1977). A California court of appeals considered two factors particularly relevant to finding a violation of the covenant in the employment context: first, the duration of the employee's employment; and, second, the extent to which the employer followed its own personnel polices in terminating the employee. Cleary v. American Airlines, 168 Cal. Rptr. 722, 729 (Ct. App. 1980). California subsequently expanded this list in Rulon-Miller v. IBM, 208 Cal. Rptr. 524 (Ct. App. 1984). There, the court listed the following factors:

> (1) whether or not the employee was discharged for legitimate business and employment reasons; (2) whether or not the employee was discharged on a pretext, that is, for a false reason or motive put forth to hide the real one; (3) whether or not the employee was engaged in a sensitive or confidential management position; (4) whether or not the employee has a conflict of interest; (5) whether or not the employee's personal, private or social relationships endangered, injured or jeopardized the employer's legitimate business interests; (6) whether or not the employer violated, invaded or infringed upon the employee's personal privacy and personal, private and social relationships; (7) whether or not the employee was discriminated against by the employer because of that employee's sex.

jurisdictions, including Utah, recognize that all contracts contain an implied covenant of good faith and fair dealing.[30] Many jurisdictions, however, have been reluctant to adopt the covenant as a substantive limitation on an employer's right to terminate at-will employees.[31] This reluctance reflects the sentiment that the scope of the covenant in the employment context is incapable of precise definition,[32] leading to unpredictable and inconsistent application.[33] Courts also are split on whether they should award contract or tort damages for a breach of the implied covenant of good faith and fair dealing.[34] As a result of these uncertainties, the implied covenant of good faith and fair dealing apparently is the least recognized of the three exceptions to the at-will rule.[35]

Other courts, however, have extended the implied covenant of good faith and fair dealing to the employment relationship.[36] In *Cleary v. American Airlines*,[37] for example, the California Court of Appeals held that terminating an employee without good cause after eighteen years of service breached the implied covenant of good faith and fair dealing.[38] The court stated that

---

Id. at 532 n.6.

30. *See, e.g.*, Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1046 (Utah 1989) (Durham & Stewart, JJ.) (citing Beck v. Farmers Ins. Exch., 701 P.2d 795, 798 (Utah 1985)) (discussing the application of implied covenant of good faith and fair dealing to employment contracts).

31. *See, e.g.*, Murphy v. American Home Prod. Corp., 448 N.E.2d 86, 91 (N.Y. 1983) (refusing to adopt covenant); Thompson v. St. Regis Paper Co., 685 P.2d 1081, 1086 (Wash. 1984) (same).

32. *See* Chagares, *supra* note 7, at 372.

33. *See Berube*, 771 P.2d at 1051-52 (Zimmerman, J., concurring).

34. *Compare* Cleary v. American Airlines, 168 Cal. Rptr. 722, 729 (Ct. App. 1980) (breach sounds in tort and contract) *with* Foley v. Interactive Data Corp., 765 P.2d 373, 398 (Cal. 1988) (breach sounds in contract only).

35. *See* Chagares, *supra* note 7, at 372.

36. The implied covenant of good faith and fair dealing has been judicially adopted, in various forms, in at least 11 jurisdictions. Hoffman-La Roche, Inc. v. Campbell, 512 So. 2d 725, 738-39 (Ala. 1987); Mitford v. De Lasala, 666 P.2d 1000, 1007 (Alaska 1983); Wagenseller v. Scottsdale Memorial Hosp., 710 P.2d 1025, 1038-41 (Ariz. 1985); Smith v. American Greetings Corp., 804 S.W.2d 683, 684 (Ark. 1991); Cleary v. American Airlines, 168 Cal. Rptr. 722, 727-28 (Ct. App. 1980); Magnan v. Anaconda Indus., Inc., 479 A.2d 781, 788 (Conn. 1984); Fortune v. National Cash Register Co., 364 N.E.2d 1251, 1256-57 (Mass. 1977); Gates v. Life of Montana Ins. Co., 638 P.2d 1063, 1067 (Mont. 1982); K Mart Corp. v. Ponsok, 732 P.2d 1364, 1372 (Nev. 1987); Monge v. Beebe Rubber Co., 316 A.2d 549, 551 (N.H. 1974); Elliott v. Tektronix, Inc., 796 P.2d 361, 365-66 (Or. Ct. App. 1990).

37. 168 Cal. Rptr. 722 (Ct. App. 1980).

38. *Id.* at 729.

the covenant applies to all contracts,[39] and is "*unconditional* and *independent* in nature[,]"[40] because it is "necessary to ensure social stability in our society."[41] Although *Cleary* permitted tort-based damages for breach of the implied covenant of good faith and fair dealing, California has since retreated from this position and now limits recovery to contract damages.[42]

2. Public Policy

Courts recognizing the public policy exception[43] allow employees to recover for wrongful discharge if they can demonstrate that their termination violates some recognized and established public policy.[44] The obvious problem with applying this exception, however, is defining "public policy."[45] Applying this exception requires courts not only to define legitimate sources of public policy, but to define the principles that support a particular policy as well. Further, courts must make qualitative judgments as to whether a policy, once established, should restrict an employer's right to terminate at-will employees.[46] In making these determinations, courts have sought guidance by looking to statutes, constitutions,[47] and judicial decisions.[48] Some courts additionally look to whether the policy advanced as limiting the employer's right to terminate substantially weighs in favor of

---

39. *Id.* at 728.

40. *Id.*

41. *Id.* at 729; *see also* Pugh v. See's Candies, Inc., 171 Cal. Rptr. 917, 927 (Ct. App. 1981) (terminating employee after 32 years of employment violates implied covenant irrespective of internal termination procedures); *National Cash Register*, 364 N.E.2d at 1256 (employer and employee must act in good faith).

42. *See supra* note 34 (demonstrating California's change to contract-based damages).

43. For the seminal article on the public policy exception to the employment-at-will doctrine, see Lawrence E. Blades, *Employment At Will vs. Individual Freedom: On Limiting the Abusive Exercise of Employer Power*, 67 COLUM. L. REV. 1404 (1967).

44. *See* Chagares, *supra* note 7, at 369.

45. *See, e.g., Petermann*, 344 P.2d at 27 ("The term 'public policy' is inherently not subject to precise definition.").

46. *See Berube*, 771 P.2d at 1042-43 (Durham & Stewart, JJ.).

47. *See* Firestone Textile Co. v. Meadows, 666 S.W.2d 730, 731 (Ky. 1983); Hodges v. Gibson Prods. Co., 811 P.2d 151, 165-67 (Utah 1991); *Berube*, 771 P.2d at 1043 (Durham & Stewart, JJ.); Brockmeyer v. Dun & Bradstreet, 335 N.W.2d 834, 840 (Wis. 1983).

48. *See* Wagenseller v. Scottsdale Memorial Hosp., 710 P.2d 1025, 1033-34 (Ariz. 1985); *Berube*, 771 P.2d at 1043 (Durham & Stewart, JJ.); Thompson v. St. Regis Paper Co., 685 P.2d 1081, 1089 (Wash. 1984).

creating an exception to the at-will rule.[49] Despite these difficulties, the public policy exception is the most widely recognized exception to the at-will rule.[50]

Decisions recognizing a public policy exception generally fall into three categories.[51] The first category involves cases where an employee is terminated "for refusing to commit an unlawful or wrongful act."[52] For example, in *Tameny v. Atlantic Richfield*

---

49. *See, e.g., Berube*, 771 P.2d at 1043 (Durham & Stewart, JJ.) (state's policy promoting job security and full employment is insufficient to maintain wrongful termination claim).

50. The public policy exception to the employment-at-will rule has been judicially recognized in at least 41 jurisdictions. Knight v. American Guard & Alert, Inc., 714 P.2d 788, 792 (Alaska 1986); Wagenseller v. Scottsdale Memorial Hosp., 710 P.2d 1025, 1033 (Ariz. 1985); Sterling Drug, Inc. v. Oxford, 743 S.W.2d 380, 383-85 (Ark. 1988); Tameny v. Atlantic Richfield Co., 610 P.2d 1330, 1336-37 (Cal. 1980); Martin Marietta Corp. v. Lorenz, 823 P.2d 100, 108 (Colo. 1992); Sheets v. Teddy's Frosted Foods, Inc., 427 A.2d 385, 388-89 (Conn. 1980); Smith v. Piezo Technology & Professional Adm'r, 427 So. 2d 182, 184 (Fla. 1983); Parnar v. Americana Hotels, Inc., 652 P.2d 625, 631 (Haw. 1982); MacNeil v. Minidoka Memorial Hosp., 701 P.2d 208, 209 (Idaho 1985); Kelsay v. Motorola, Inc., 384 N.E.2d 353, 357 (Ill. 1978); Frampton v. Central Ind. Gas Co., 297 N.E.2d 425, 428 (Ind. 1973); Springer v. Wecks & Leo Co., 429 N.W.2d 558, 560 (Iowa 1988); Anco Constr. Co. v. Freeman, 693 P.2d 1183, 1186 (Kan. 1985); Firestone Textile Co. v. Meadows, 666 S.W.2d 730, 732-33 (Ky. 1983); Wiley v. Missouri Pac. R.R., 430 So. 2d 1016, 1022 (La. Ct. App. 1982); Adler v. American Standard Corp., 432 A.2d 464, 471 (Md. 1981); DeRose v. Putnam Management Co., 496 N.E.2d 428, 431 (Mass. 1986); Suchodolski v. Michigan Consol. Gas Co., 316 N.W.2d 710, 711-12 (Mich. 1982); Phipps v. Clark Oil & Ref. Corp., 408 N.W.2d 569, 571 (Minn. 1987); Laws v. Aetna Fin. Co., 667 F. Supp. 342, 346-48 (N.D. Miss. 1987) (applying Mississippi law); Beasley v. Affiliated Hosp. Prods., 713 S.W.2d 557, 560 (Mo. Ct. App. 1986); Keneally v. Orgain, 606 P.2d 127, 129 (Mont. 1980); Ambroz v. Cornhusker Square Ltd., 416 N.W.2d 510, 514-15 (Neb. 1987); Hansen v. Harrah's, 675 P.2d 394, 396 (Nev. 1984); Cloutier v. Great Atl. & Pac. Tea Co., 436 A.2d 1140, 1142-43 (N.H. 1981); Pierce v. Ortho Pharmaceutical Corp., 417 A.2d 505, 512 (N.J. 1980); Boudar v. E G & G, Inc., 730 P.2d 454, 456 (N.M. 1986); Sides v. Duke Hosp., 328 S.E.2d 818, 826 (N.C. Ct. App. 1985); Krein v. Marian Manor Nursing Home, 415 N.W.2d 793, 794-95 (N.D. 1987); Burk v. K-Mart Corp., 770 P.2d 24, 28 (Okla. 1989); Nees v. Hocks, 536 P.2d 512, 515-16 (Or. 1975); Darlington v. GE, 504 A.2d 306, 317 (Pa. Super. Ct. 1986); Cummins v. EG & G Sealol, Inc., 690 F. Supp. 134, 137-39 (D.R.I. 1988) (applying Rhode Island law); Ludwick v. This Minute of Carolina, Inc., 337 S.E.2d 213, 216 (S.C. 1985); Clanton v. Cain-Sloan Co., 677 S.W.2d 441, 444-45 (Tenn. 1984); Sabine Pilot Serv., Inc. v. Hauck, 687 S.W.2d 733, 735 (Tex. 1985); Peterson v. Browning, 832 P.2d 1280, 1281, 1285-86 (Utah 1992); Payne v. Rozendaal, 520 A.2d 586, 588-90 (Vt. 1986); Bowman v. State Bank, 331 S.E.2d 797, 801 (Va. 1985); Thompson v. St. Regis Paper Co., 685 P.2d 1081, 1089 (Wash. 1984); Harless v. First Nat'l Bank, 246 S.E.2d 270, 275 (W. Va. 1978); Brockmeyer v. Dun & Bradstreet, 335 N.W.2d 834, 840 (Wis. 1983).

51. Peterson v. Browning, 832 P.2d 1280, 1281 (Utah 1992) (Durham & Stewart, JJ.); Kenneth T. Lopatka, *The Emerging Law of Wrongful Discharge—A Quadrennial Assessment of the Labor Law Issue of the 80s*, 40 BUS. LAW. 1, 6-7 (1984).

52. Lopatka, *supra* note 51, at 7.

Co.,[53] an employer terminated an employee for refusing to participate in an illegal price-fixing scheme. The court held that employers who demand their employees to commit criminal acts violate the public policy exception to the at-will rule.[54] According to the court, "[a]n employer engaging in such conduct violates a basic duty imposed by law upon all employers."[55] The *Tameny* court found this basic duty in fundamental public policies expressed in the state's criminal statutes.[56]

The second category of public policy exceptions to the at-will rule consists of cases in which an employee is terminated for performing a public obligation. In *Nees v. Hocks*,[57] for example, the Oregon Supreme Court upheld an employee's wrongful termination claim where the employer terminated the employee for serving jury duty.[58] Although the court agreed that Oregon followed the traditional at-will rule, it reasoned that in some circumstances terminating an employee interferes with important community interests.[59] The *Hocks* court stated that jury duty is an important civil obligation and that the good of the community would be affected adversely if employers could terminate employees for engaging in it.[60] Therefore, the court held, employees terminated for serving jury duty can sue their former employers for wrongful discharge.[61]

The final category of terminations that violate public policy involve cases in which employers terminate employees for exercising legal rights or privileges.[62] This category includes termination for pursuing a workers' compensation claim,[63] for

---

53. 610 P.2d 1330 (Cal. 1980).

54. *Id.* at 1336-37; *see also* Petermann v. International Bhd. of Teamsters Local 396, 344 P.2d 25, 27 (Cal. Dist. Ct. App. 1959) (termination for refusing to commit perjury violates public policy).

55. *Tameny*, 610 P.2d at 1337.

56. *Id.* at 1335.

57. 536 P.2d 512 (Or. 1975).

58. *Id.* at 516; *see also* Reuther v. Fowler & Williams, Inc., 386 A.2d 119, 121-22 (Pa. Super. Ct. 1978) (recognizing wrongful termination cause of action when employee was fired for serving jury duty).

59. *Hocks*, 536 P.2d at 515.

60. *Id.* at 516.

61. *Id.*

62. *See* Lopatka, *supra* note 51, at 11.

63. *See, e.g.*, Kelsay v. Motorola, Inc., 384 N.E.2d 353, 357 (Ill. 1978) (employer's power to terminate should not prevail when doing so prevents employees from pursuing workers' compensation claim); Frampton v. Central Ind. Gas Co., 297 N.E.2d 425, 428

refusing to take a polygraph test, and for refusing to submit to a serious invasion of privacy.[64]

3. *Implied-In-Fact Contract*

The at-will rule traditionally was bolstered by the contract doctrines of additional consideration and mutuality of obligation.[65] According to the requirement of additional consideration, employees must provide some form of consideration in addition to their continued employment to support a contract for employment.[66] Because employees can quit at any time, their relationship with their employers similarly lacked the requisite mutuality of obligation to support a contract for employment.[67] Nonetheless, some courts have been willing to find implied contracts[68] between

---

(Ind. 1973) (same); Firestone Textile Co. v. Meadows, 666 S.W.2d 730, 732 (Ky. 1984) (same).

64. *See* Lopatka, *supra* note 51, at 11 & nn.57-58.

65. *See id.* at 17-18; Kelly McWilliams, Note, *The Employment Handbook as a Contractual Limitation on the Employment At Will Doctrine*, 31 VILL. L. REV. 335, 351 (1986).

66. *See* Lopatka, *supra* note 51, at 17-18.

67. *See* McWilliams, *supra* note 65, at 352.

68. At least 41 jurisdictions have judicially adopted an implied contract exception to the at-will rule. Hoffmann-La Roche, Inc. v. Campbell, 512 So. 2d 725, 733-35 (Ala. 1987); Eales v. Tanana Valley Medical-Surgical Group, Inc., 663 P.2d 958, 959-60 (Alaska 1983); Leikvold v. Valley View Community Hosp., 688 P.2d 170, 172-74 (Ariz. 1984); Gladden v. Arkansas Children's Hosp., 728 S.W.2d 501, 505 (Ark. 1987); Pugh v. See's Candies, Inc., 171 Cal. Rptr. 917, 922 (Ct. App. 1981); Continental Air Lines v. Keenan, 731 P.2d 708, 711-12 (Colo. 1987); Finley v. Aetna Life & Casualty Co., 520 A.2d 208, 213-14 (Conn. 1987); Washington Welfare Ass'n v. Wheeler, 496 A.2d 613, 615-16 (D.C. 1985); Kinoshita v. Canadian Pac. Airlines, 724 P.2d 110, 115-17 (Haw. 1986); Harkness v. City of Burley, 715 P.2d 1283, 1286-88 (Idaho 1986); Duldulao v. Saint Mary of Nazareth Hosp. Ctr., 505 N.E.2d 314, 318 (Ill. 1987); Cannon v. National By-Prods., Inc., 422 N.W.2d 638, 640 (Iowa 1988); Morriss v. Coleman Co., 738 P.2d 841, 848-49 (Kan. 1987); Shah v. American Synthetic Rubber Corp., 655 S.W.2d 489, 490-92 (Ky. 1983); Larrabee v. Penobscot Frozen Foods, Inc., 486 A.2d 97, 99-100 (Me. 1984); Staggs v. Blue Cross, 486 A.2d 798, 803-04 (Md. Ct. Spec. App. 1985); Hobson v. McLean Hosp. Corp., 522 N.E.2d 975, 977 (Mass. 1988); Toussaint v. Blue Cross & Blue Shield, 292 N.W.2d 880, 885 (Mich. 1980); Pine River State Bank v. Mettille, 333 N.W.2d 622, 627-30 (Minn. 1983); Robinson v. Board of Trustees, 477 So. 2d 1352, 1353-54 (Miss. 1985); Johnston v. Pan Handle Coop. Ass'n, 408 N.W.2d 261, 266 (Neb. 1987); Southwest Gas Corp. v. Ahmad, 668 P.2d 261, 261 (Nev. 1983); Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1264-66, *modified*, 499 A.2d 515 (N.J. 1985); Forrester v. Parker, 606 P.2d 191, 192 (N.M. 1980); Weiner v. McGraw-Hill, Inc., 443 N.E.2d 441, 444-46 (N.Y. 1982); Hammond v. State Personnel Bd., 345 N.W.2d 359, 361 (N.D. 1984); Mers v. Dispatch Printing Co., 483 N.E.2d 150, 154-55 (Ohio 1985); Langdon v. Saga Corp., 569 P.2d 524, 526-28 (Okla. Ct. App. 1976); Yartzoff v. Democrat-Herald Publishing Co., 576 P.2d 356, 360 (Or. 1978);

employers and employees that limit employers' absolute right to terminate employees.[69] Such implied contracts are based on promissory statements contained in employment handbooks,[70] oral assurances,[71] or promises inferred from the conduct of the parties.[72] These promissory statements generally are either promises to terminate for "just cause" only[73] or promises to follow certain administrative procedures prior to termination.[74]

Some of the courts that forged the implied contract exception did so on the basis of general notions of equity. In *Toussaint v. Blue Cross & Blue Shield*,[75] for example, the employer gave its employees a handbook describing various termination procedures and promised "to treat employees leaving [the employer] in a fair and consistent manner and to release employees for just cause only."[76] After being fired without cause, an employee sued the employer for wrongful termination.[77]

The Michigan Supreme Court reasoned that the at-will rule was a rule of construction, rather than a rule of substantive law.[78] Moreover, the court found that employer statements of policy, such as those contained in employee handbooks, can give

---

DiBonaventura v. Consolidated Rail Corp., 539 A.2d 865, 867 (Pa. Super. Ct. 1988); Small v. Springs Indus., Inc., 357 S.E.2d 452, 454 (S.C. 1987); Osterkamp v. Alkota Mfg., Inc., 332 N.W.2d 275, 277 (S.D. 1983); Hamby v. Genesco, Inc., 627 S.W.2d 373, 375-76 (Tenn. Ct. App. 1981); Aiello v. United Air Lines, 818 F.2d 1196, 1200-01 (5th Cir. 1987) (applying Texas law); Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1045-46, 1052 (Utah 1989); Sherman v. Rutland Hosp., Inc., 500 A.2d 230, 232 (Vt. 1985); Thompson v. American Motor Inns, Inc., 623 F. Supp. 409, 416 (W.D. Va. 1985) (applying Virginia law); Thompson v. Saint Regis Paper Co., 685 P.2d 1081, 1084 (Wash. 1984); Cook v. Heck's Inc., 342 S.E.2d 453, 457-59 (W. Va. 1986); Ferraro v. Koelsch, 368 N.W.2d 666, 668 (Wis. 1985); Mobile Coal Producing, Inc. v. Parks, 704 P.2d 702, 706 (Wyo. 1985).

69. For an in-depth discussion of the implied contract exception to employment-at-will, see Shapiro & Tune, *supra* note 12.

70. *See, e.g., Toussaint*, 292 N.W.2d at 885 (handbook may become part of employment contract thereby limiting employer's right to terminate at will); *Woolley*, 491 A.2d at 1264-66 (same); *Berube*, 771 P.2d at 1044 (Durham & Stewart, JJ.) (same).

71. *See, e.g., Berube*, 771 P.2d at 1044 (Durham & Stewart, JJ.) (evidence of oral agreement may overcome presumption that employment is at will).

72. Circumstances surrounding the employer-employee relationship may overcome the presumption that employment is at will. *See id.* (Durham & Stewart, JJ.); *infra* notes 264-72 and accompanying text (discussing effect of employer's de facto policies on employment relationship).

73. *See* McWilliams, *supra* note 65, at 354.

74. *See id.*; Chagares, *supra* note 7, at 373 & n.66.

75. 292 N.W.2d 880 (Mich. 1980).

76. *Id.* at 893.

77. *Id.* at 883.

78. *Id.* at 885.

rise to contractual rights.[79] This is true even though "the statement of policy is signed by neither party, can be unilaterally amended by the employer without notice to the employee, and contains no reference to a specific employee, his job description or compensation."[80]

The court did not delve into the specifics of these contractual rights, but simply reasoned that employer statements ensuring job security enabled the employer to "secure[] an orderly, cooperative and loyal work force."[81] Such cooperation benefitted the employer.[82] The court, therefore, interpreted the handbook consistent with the employee's legitimate expectations.[83]

The *Toussaint* court also recognized that employers were free to require employees to acknowledge that their employment was at will.[84] Where an employer established employment policies and made those policies known to its employees, however, the employer "created a situation 'instinct with an obligation.'"[85]

Other courts based the implied contract exception on traditional contract principles. In *Woolley v. Hoffmann-La Roche, Inc.*,[86] for example, the employee received a handbook from his employer after commencement of employment.[87] The handbook specified a number of grounds for termination and procedures the employer would follow in the event of termination.[88] The employee claimed that the employer's failure to terminate him for one of the specified reasons, and failure to follow the prescribed procedures, breached an implied-in-fact contract.[89]

In holding for the employee, the New Jersey Supreme Court agreed with the general principles announced in *Toussaint*.[90] According to the court, the mere distribution of a handbook

---

79. *Id.*
80. *Id.* at 892.
81. *Id.*
82. *Id.*
83. *See id.* at 893.
84. *Id.* at 890-91.
85. *Id.* at 892 (quoting Wood v. Lucy, Lady Duff-Gordon, 118 N.E. 214, 214 (N.Y. 1917)).
86. 491 A.2d 1257, *modified*, 499 A.2d 515 (N.J. 1985).
87. *Id.* at 1258.
88. *Id.* at 1271-73.
89. *Id.* at 1258.
90. *Id.* at 1267-68.

suggested its importance.[91] Because the handbook contained provisions regarding job security, something of paramount importance to employees, the employees could reasonably conclude that the handbook created a legally binding obligation.[92] From this the court reasoned that handbooks constitute an employer's offer for a unilateral contract,[93] and employees accept the unilateral offer by continuing to work.[94] The court found valid consideration for this contract in the fact that employees continue to work even though not legally obligated to do so.[95] The unilateral contract analysis announced in *Woolley* became the paradigm for jurisdictions that follow the traditional contract analysis in formulating the implied-in-fact contract exception to at-will employment.[96]

The recognition of the exceptions outlined above signaled an erosion of the at-will rule.[97] The cases that began this erosion initially were viewed as isolated aberrations from the nearly uniform application of the at-will rule.[98] The body of law that developed from these cases, however, represents one of the most significant protections for non-union, private-sector employees against arbitrary or malicious discharge.[99]

---

91. *Id.* at 1265. Or as another judge put it, handbooks are not "corporate illusion[s], 'full of sound . . . signifying nothing.'" Weiner v. McGraw-Hill, Inc., 442 N.Y.S.2d 11, 11-12 (1981) (Kupferman, J., dissenting).

92. *Woolley*, 491 A.2d at 1265-66. Importantly, the court focused on the employee's reasonable expectations rather than the employer's subjective intent: "Whatever Hoffmann-La Roche may have intended, that which was read by its employees was a promise not to fire them except for cause." *Id.* at 1266.

93. *Id.* at 1264-66. In a bilateral contract, offerors bargain for a promise to perform. *Id.* at 1267. In a unilateral contract, however, offerors bargain for some act on the part of the offeree that, if given, results in a binding obligation. *Id.*

94. *Id.*

95. *Id.* The court skirted the mutuality of obligation issue by claiming that although it raised potential interpretative problems, it did not make the job security provisions unenforceable. *Id.* at 1269.

96. *See, e.g., Johnson*, 818 P.2d at 1001-02 (indicating Utah and other jurisdictions follow unilateral contract approach).

97. *See* Peck, *supra* note 14, at 732.

98. *See supra* notes 22-27 and accompanying text (discussing *Petermann* decision).

99. *See generally* Lopatka, *supra* note 51, at 2-3 (discussing impact of exceptions to at-will rule).

C. The At-Will Rule in Utah

Development of the employment-at-will rule in Utah generally followed its development in other jurisdictions. The Utah Supreme Court first considered the at-will issue in *Price v. Western Loan & Savings Co.*[100] *Price* involved a written employment contract that guaranteed the employee would not be terminated without good reason, so long as his services were necessary.[101] The court applied traditional rules of contract construction and concluded that because the contract did not specify an exact period of employment, "it lacked the essential element of mutuality of obligation and was terminable at will by either party."[102] Similarly, in *Bullock v. Deseret Dodge Truck Center, Inc.*,[103] the employee, relying on the employer's representations of permanent employment, moved from Texas to Utah to become a sales manager.[104] The employee and employer entered into a written contract with detailed stock ownership and transfer provisions.[105] After being terminated, the employee sued for wrongful discharge, claiming there was an implicit agreement that his employer would employ him for eight years.[106] The court, however, refused to infer a period of employment.[107] Because the employee could not establish that the employment contract was for a definite period, the court held that the employer could terminate the employee at will.[108] Thus, Utah strictly applied the at-will rule and deflected contractual challenges by invoking the mutuality of obligation requirement.

---

100. 100 P. 677 (Utah 1909).

101. *Id.* at 678.

102. *Id.* at 680; *see also* Hancock v. Luke, 173 P. 137, 140 (Utah 1918) (where employment contract did not specify duration of employment either party could terminate at will).

103. 354 P.2d 559 (Utah 1960).

104. *Id.* at 560-61.

105. *Id.* at 560 & n.1.

106. *Id.* at 560-61.

107. *Id.* at 562.

108. *Id.*; *see also* Bihlmaier v. Carson, 603 P.2d 790, 792 (Utah 1979) (employee hired for indefinite term has no right of action against employer for breach of employment contract); Crane Co. v. Dahle, 576 P.2d 870, 872 (Utah 1978) ("In the absence of a contract for a definite term, an employee may quit whenever he desires, the same as the employer may fire him.").

The Utah Supreme Court first discussed implied employment contracts in *Held v. American Linen Supply Co.*[109] In *Held*, the court stated that whether an employee has a right to be terminated for cause depends on the terms of the contract, whether "express or implied."[110] The court found no provision in the employee's collective bargaining agreement from which to infer such a right and consequently held that she could be terminated at will.[111] Utah first recognized an implied employment contract in *Piacitelli v. Southern Utah State College.*[112] The employee in *Piacitelli* claimed that a personnel manual created an implied-in-fact contract. This contract, the employee argued, prevented his termination unless the employer complied with the termination procedures outlined in the manual.[113] The Utah Supreme Court accepted the trial court's conclusion that the employer breached its implied contract with the employee when it failed to follow those termination procedures.[114]

In *Rose v. Allied Development Co.*,[115] decided in 1986, the court observed that Utah recognized only two limited exceptions to the at-will rule: first, where the employee offers consideration, in addition to regular services, to support the employer's promise of job security;[116] and second, where the employer and employee enter an express or implied agreement regarding the duration of the employee's employment.[117]

In 1989, the Utah Supreme Court significantly expanded the implied contract exception in *Berube v. Fashion Centre, Ltd.*[118] The employee in *Berube* argued that a written disciplinary policy

---

109. 307 P.2d 210 (Utah 1957).

110. *Id.* at 211.

111. *Id.* at 211-12.

112. 636 P.2d 1063 (Utah 1981).

113. *Id.* at 1064. The employment manual at issue specified two types of employees; probationary, which could be terminated at the will of the employer, and permanent, which could be terminated only after the employer followed certain procedures set forth in the manual. *Id.* The trial court determined that the employee at issue was "permanent." *Id.* at 1064-65.

114. *Id.* at 1070. The court stated that an employee manual could give rise to contractual obligations on the part of the employer, even though the employer could unilaterally amend the manual without notice to the employee. *Id.* at 1066 n.5 (citing Toussaint v. Blue Cross & Blue Shield, 292 N.W.2d 880, 892 (Mich. 1980)).

115. 719 P.2d 83 (Utah 1986).

116. *Id.* at 85 (citing Bihlmaier v. Carson, 603 P.2d 790, 792 (Utah 1979)).

117. *Id.*

118. 771 P.2d 1033 (Utah 1989).

created an implied-in-fact contract requiring her employer to terminate for just cause only.[119] The lead opinion[120] in *Berube* reviewed the history of the at-will doctrine[121] and outlined its three general exceptions. These exceptions, and subsequent cases considering them, are discussed separately below.

1. Implied Covenant of Good Faith and Fair Dealing

Utah does not presently recognize an exception to the at-will rule based on the implied covenant of good faith and fair dealing.[122] In *Berube*, only Justices Durham and Stewart expressed a willingness to recognize this exception, and if recognized, would award contract rather than tort damages where it applied.[123] Justices Howe and Hall believed it unnecessary to consider the implied covenant in *Berube* because the issue was not properly before the court.[124] Justice Zimmerman criticized Justice Durham's endorsement of the implied covenant of good faith and fair dealing exception because it failed to "establish predictable guidelines for determining what [the] duty is and when an employer can be found to owe such a duty to an employee."[125] Although Justice Zimmerman hinted that he might be willing to consider the implied covenant of good faith and fair dealing should the appropriate case present itself,[126] the court's pronouncements since *Berube* make it unlikely that the exception will be recognized in Utah in the near future.[127]

---

119. *Id.* at 1048-49.

120. The court's opinion was divided into three opinions. Justice Durham authored the lead opinion with Justice Stewart concurring. *Id.* at 1050. Justice Howe wrote a separate concurrence and was joined by Justice Hall. *Id.* Finally, Justice Zimmerman wrote a separate opinion, concurring in the result. *Id.*

121. *Id.* at 1040-41 (Durham & Stewart, JJ.).

122. *See* Heslop v. Bank of Utah, 194 Utah Adv. Rep. 20, 26 (Sept. 4, 1992); Brehany v. Nordstrom, Inc., 812 P.2d 49, 55 (Utah 1991); Loose v. Nature-All Corp., 785 P.2d 1096, 1098 (Utah 1989).

123. *See Berube*, 771 P.2d at 1046-47 (Durham & Stewart, JJ.).

124. *Id.* at 1050 (Howe, J., & Hall, C.J., concurring). Justice Howe did not think it was necessary or appropriate to go beyond the handbook, which he viewed as "part of the total employment contract." *Id.* (Howe, J., & Hall, C.J., concurring).

125. *Id.* at 1051-52 (Zimmerman, J., concurring).

126. *Id.* (Zimmerman, J., concurring). Justice Zimmerman stated, "[u]ntil we have had a better opportunity to consider the minimum rights and obligations that inhere in the employment relationship . . . I would reject invitations to create this cause of action." *Id.* (Zimmerman, J., concurring).

127. *See, e.g.*, Heslop v. Bank of Utah, 194 Utah Adv. Rep. 20, 26-27 (Sept. 4, 1992)

2. Public Policy

Utah currently recognizes a public policy exception to the employment-at-will rule.[128] The precise scope of the exception, however, remains undefined.[129] In *Peterson v. Browning*,[130] the Utah Supreme Court unanimously recognized a public policy exception.[131] Over a strong dissent, the court also held that violations sound in tort, rather than contract.[132] The court further noted that to violate the public policy exception, the employer's conduct must violate some clear expression of the public conscience, and society's affected interests must be substantial.[133]

(refusing to recognize implied covenant of good faith and fair dealing in employment context); Brehany v. Nordstrom, Inc., 812 P.2d 49, 55-56 (Utah 1991) (covenant of good faith and fair dealing cannot be implied where no express limit on employers' power to terminate); Loose v. Nature-All Corp., 785 P.2d 1096, 1098 (Utah 1989) (refusing to recognize implied covenant of good faith and fair dealing in employment context); Caldwell v. Ford, Bacon & Davis Utah, Inc., 777 P.2d 483, 485 (Utah 1989) (same).

128. *See Heslop*, 194 Utah Adv. Rep. at 27; Peterson v. Browning, 832 P.2d 1280, 1281 (Utah 1992) (Durham & Stewart, JJ.); *id.* at 1285-86 (Howe, J., concurring); *id.* at 1286 (Zimmerman, J., & Hall, C.J., dissenting); *see also Berube*, 771 P.2d at 1043 (Durham & Stewart, JJ.); *id.* at 1051 (Zimmerman, J., concurring). In *Berube*, Justice Zimmerman agreed that Utah should recognize a public policy exception to the at-will rule, but he was not "prepared to say what the precise content of that exception should be." *Id.* (Zimmerman, J., concurring).

129. *Peterson*, 832 P.2d at 1282 (Durham & Stewart, JJ.) ("The identification of clear and substantial public policies will require case-by-case development."); *see also id.* at 1287-88 (Zimmerman, J., & Hall, C.J., dissenting) (criticizing majority's failure to provide guidance to future litigants).

130. 832 P.2d 1280 (Utah 1992). The United States District Court for the District of Utah certified two questions to the Utah Supreme Court for resolution. First, whether the public policy exception to the employment-at-will rule encompasses violations of federal law and the law of other states. Second, whether the public policy exception sounds in tort or contract. *Id.* at 1281 (Durham & Stewart, JJ.).

131. *Id.* at 1281 (Durham & Stewart, JJ.); *id.* at 1285-86 (Howe, J., concurring); *id.* at 1286 (Zimmerman, J., & Hall, C.J., dissenting). Although Justice Zimmerman and Chief Justice Hall dissented, both agreed that discharges in violation of public policy are actionable. *Id.* at 1286 (Zimmerman, J., & Hall, C.J., dissenting).

In a subsequent case, the Utah Supreme Court announced a three-part inquiry for determining violations of the public policy exception: "(1) Does the plaintiff's termination implicate a clear and substantial public policy? (2) Did the employer violate this public policy by requiring the employee to engage in conduct violating the policy or by punishing conduct furthering the policy? (3) Was violation of this public policy a substantial factor in the plaintiff's termination?" *Heslop*, 194 Utah Adv. Rep. at 25.

132. *Peterson*, 832 P.2d at 1285 (Durham & Stewart, JJ.); *id.* at 1285-86 (Howe, J., concurring). *But see id.* at 1288-94 (Zimmerman, J., & Hall, C.J., dissenting).

133. *Id.* at 1283 (Durham & Stewart, JJ.). Utah's public policies, however, are not necessarily confined to Utah law. The *Peterson* court held that violations of federal law

In *Berube*, Justice Durham identified several sources for public policy exceptions, including legislative enactments[134] and judicial decisions.[135] In discussing the scope of the public policy exception, Justice Durham cautioned that the exception should be construed narrowly and policies relied on must be "fundamental and permanent" rather than "superficial and transitory."[136] Thus, the public policy exception cannot be read so broadly as to establish a requirement that termination be for cause only.[137]

Justice Howe echoed this concern in *Peterson*. In a concurring opinion, he emphasized that the public policy exception should be narrowly construed and infrequently invoked.[138] The *Peterson* dissent agreed that the public policy must in fact be clear and substantial to give rise to an actionable discharge.[139] However, according to the dissent, the clear and substantial *test* fails to give adequate guidance to the bench and bar.[140]

The *Peterson* court was sharply divided over whether a court should award contract or tort damages for violations of the public

---

or the law of other states may properly form the basis of a wrongful discharge claim. *Id.* at 1283 (Durham & Stewart, JJ.); *id.* at 1285-86 (Howe, J., concurring); *id.* at 1288 (Zimmerman, J., & Hall, C.J., dissenting). The court carefully limited its holding, however, to violations of federal or state law that contravene "clear and substantial public policy of the state of Utah." *Id.* at 1283 (Durham & Stewart, JJ.). Moreover, the plaintiff must establish a connection between the law violated and the public policies of Utah. *Id.* (Durham & Stewart, JJ.)

134. *Berube*, 771 P.2d at 1043 (Durham & Stewart, JJ.). An employee can only rely on legislative enactments that "protect the public or promote public interest" for exceptions to the at-will rule based on public policy. *Id.* (Durham & Stewart, JJ.).

135. *Id.* (Durham & Stewart, JJ.). Employees can only rely on judicial decisions that announce public policy in areas not addressed by the legislature. *Id.* (Durham & Stewart, JJ.).

136. *Id.* (Durham & Stewart, JJ.).

137. Caldwell v. Ford, Bacon & Davis Utah, Inc., 777 P.2d 483, 485 & n.2 (Utah 1989).

138. *Peterson*, 832 P.2d at 1285 (Howe, J., concurring). According to Justice Howe, "'[t]he public policy exception is narrow enough in its scope and application to be no threat to employers who operate within the mandates of the law and clearly established public policy as set out in the duly adopted laws.'" *Id.* (Howe, J., concurring) (quoting Boyle v. Vista Eyewear, Inc., 700 S.W.2d 859, 878 (Mo. Ct. App. 1985)).

139. *Id.* at 1288 (Zimmerman, J., & Hall, C.J., dissenting).

140. *Id.* (Zimmerman, J., & Hall, C.J., dissenting). Justice Zimmerman attempted to clarify the clear and substantial test by posing two questions to help give meaning to the term "substantial." First, "[i]s the policy in question one that is of sufficient importance to the public, as opposed to the parties only, that it should constitute an uncompromising bar to discharge?" *Id.* (Zimmerman, J., & Hall, C.J., dissenting). Second, "[i]s it a policy that a court would permit the parties to derogate by express contract?" *Id.* (Zimmerman, J., & Hall, C.J., dissenting).

policy exception. A majority of the court held that the public policy exception sounds in tort.[141] According to the lead opinion, "potential punitive damages will exert a valuable deterrent effect on employers who might otherwise subject their employees to a choice between violating the law and losing their jobs."[142] The dissent disagreed, arguing that contract damages are ordinarily sufficient to make the employee whole.[143] The dissent would allow discharged employees to recover punitive damages only if the employee could prove an independent tort.[144] This two-layered approach, the dissent argued, would provide the deterrent effect of the tort approach while avoiding the negative consequences of a vague, ill-defined tort remedy.[145]

Thus, in *Peterson*, the Utah Supreme Court adopted the public policy exception and identified the standard that terminated employees must meet to invoke the exception.[146] The court did not, however, attempt to further define the standard, concluding that the clear and substantial test required case-by-case development.[147] That development clearly is still in its infancy.[148]

---

141. *Id.* at 1284 (Durham & Stewart, JJ.); *id.* at 1285-86 (Howe, J., concurring).

142. *Id.* at 1285 (Durham & Stewart, JJ.).

143. *Id.* at 1288 (Zimmerman, J., & Hall, C.J., dissenting).

144. *Id.* (Zimmerman, J., & Hall, C.J., dissenting).

145. *Id.* at 1291 (Zimmerman, J., & Hall, C.J., dissenting); *see also* Beck v. Farmers Ins. Exch., 701 P.2d 795, 802 (Utah 1985) (adopting two-layered damages approach).

146. *See supra* notes 130-45 and accompanying text (discussing *Peterson* decision).

147. *Peterson*, 832 P.2d at 1287-88 (Durham & Stewart, JJ.).

148. Hodges v. Gibson Products Co., 811 P.2d 151 (Utah 1991), was the first Utah case to directly consider the public policy exception. In *Hodges*, Justices Stewart and Durham concluded that an employer's false accusations of theft and improper filing of criminal charges violated public policy. *Id.* at 167 (Stewart & Durham, JJ.). This opinion, however, did not command a majority of the court. The remaining Justices refused to join in the lead opinion's discussion of the public policy exception. *Id.* at 168 (Howe, J., concurring); *id.* (Zimmerman, J., & Hall, C.J., concurring).

In *Peterson*, a majority of the court held that an employer's attempt to force its employee to violate Missouri state tax law and federal customs statutes violated a clear and substantial public policy of the state of Utah. *Peterson*, 832 P.2d at 1283 (Durham & Stewart, JJ); *id.* at 1285-86 (Howe, J., concurring); *id.* at 1288 (Zimmerman, J., & Hall, C.J., dissenting). Although Justice Zimmerman and Chief Justice Hall dissented, both agreed that the alleged conduct violated Utah's public policy. *Id.* (Zimmerman, J., & Hall, C.J., dissenting).

Shortly after *Peterson*, the court decided Heslop v. Bank of Utah, 194 Utah Adv. Rep. 20 (Sept. 4, 1992). In *Heslop*, the court held that terminating an employee for insisting that the employer comply with state banking regulations violates public policy. *Id.* at 25. According to the court, state banking regulations "[do] not merely regulate the relationship between private individuals such as employer and employee." *Id.* Rather, they "[serve] a substantial public policy because [they] protect the public as well as

3. *Implied-In-Fact Contract*

The *Berube* court unanimously agreed that an implied-in-fact contract could modify at-will employment.[149] Consistent with *Toussaint*, in *Berube* the Utah Supreme Court reasoned that the at-will rule is simply a rule of contract construction rather than substantive law.[150] Therefore, the rule creates only a presumption that an indefinite hiring may be terminated at any time by either party.[151] The presumption "can be overcome by an affirmative showing by the plaintiff that the parties expressly or impliedly intended a specified term or agreed to terminate the relationship for cause alone."[152] The lead opinion listed potential sources for implied-in-fact contracts, including employment handbooks,[153] oral agreements,[154] and "all circumstances of the relationship which demonstrate the intent to terminate only for cause or to continue employment for a specified period."[155]

Berube held that employees have the burden of establishing the existence of an implied-in-fact contract.[156] Moreover, the lead opinion indicated that employment contracts should be construed to give effect to the parties' intentions,[157] and may not contradict written contract terms.[158] "The determination of whether sufficient indicia of an implied-in-fact promise exists is a question of fact for the jury . . . ."[159]

---

regulate[] the institutions themselves." *Id.*

149. *Berube*, 771 P.2d at 1044-45 (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J., concurring); *id.* at 1052-53 (Zimmerman, J., concurring).

150. *Id.* at 1044 (Durham & Stewart, JJ.).

151. *Id.* (Durham & Stewart, JJ.).

152. *Id.* (Durham & Stewart, JJ.).

153. *Id.* (Durham & Stewart, JJ.). The court unanimously agreed that employee handbooks were a proper source of implied contract terms. *Id.* (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J., concurring); *id.* at 1052 (Zimmerman, J., concurring).

154. *Id.* at 1044 (Durham & Stewart, JJ.). Only the lead opinion recognized oral agreements. *Id.* (Durham & Stewart, JJ.). Justice Howe deemed it "not necessary to go beyond the written policy manual of the employer." *Id.* at 1050 (Howe, J., & Hall, C.J., concurring). Justice Zimmerman identified "representations made by the employer in employee manuals, bulletins, and the like" as appropriate sources for implied contracts. *Id.* at 1052 (Zimmerman, J., concurring).

155. *Id.* at 1044 (Durham & Stewart, JJ.).

156. *Id.* (Durham & Stewart, JJ.); *id.* at 1052 (Zimmerman, J., concurring).

157. *Id.* at 1044 (Durham & Stewart, JJ.).

158. *Id.* (Durham & Stewart, JJ.).

159. *Id.* (Durham & Stewart, JJ.).

Following *Berube*, independent consideration is no longer required to support an implied contract.[160] Rather, an employee's continued employment after the employer's promissory representation is deemed sufficient consideration.[161] Similarly, the court abandoned the mutuality of obligation requirement.[162] The employee's right to terminate the employment relationship at any time does not give rise to a concomitant right in the employer. According to the court, the employee's freedom to quit in a modern economy is largely illusory.[163] Thus, mutuality of obligation cannot be invoked to avoid implied-in-fact contract liability for breach of an employer's promissory statements.

Though some courts have ruled against implied contract claims as a matter of law,[164] others have found that the rule announced in *Berube* gives rise to a question of fact, thereby precluding summary judgment.[165] The implied contract exception therefore has proven successful in mitigating the harshness of the at-will rule. Unfortunately, the Utah Supreme Court's recent pronouncement in *Johnson v. Morton Thiokol, Inc.*[166] raises serious doubts about its continued ability to do so.

---

160. *Id.* at 1045 (Durham & Stewart, JJ.).

161. *Id.* (Durham & Stewart, JJ.) (citing Yartzoff v. Democrat-Harold Publishing Co., 576 P.2d 356, 359 (Or. 1978)).

162. *Id.* (Durham & Stewart, JJ.).

163. *Id.* (Durham & Stewart, JJ.).

164. *See* Johnson v. Morton Thiokol, Inc., 818 P.2d 997, 1002-04 (Utah 1991) (disclaimer prohibited formation of implied employment contract as matter of law); *Brehany*, 812 P.2d at 56-57 (as matter of law employer may terminate employees who admit violating handbook provisions where employment manual so provides); Caldwell v. Ford, Bacon & Davis Utah, Inc., 777 P.2d 483, 485-86 (Utah 1989) (no wrongful termination where employer complied with handbook provisions when terminating employee); *see also* Arnold v. B.J. Titan Serv. Co., 783 P.2d 541, 544 (Utah 1989) (entering judgment for employee where trial court found violation of handbook procedures).

165. *See, e.g.*, Howcroft v. Mountain States Tel. & Tel. Co., 712 F. Supp. 1514, 1519-20 (D. Utah 1989) (concluding that "whether sufficient indicia of an implied-in-fact promise exists is a question of fact for the jury"); Brehany v. Nordstrom, Inc., 812 P.2d 49, 57 (Utah 1991) ("[T]riable factual issues exist as to whether the dismissal rules in the manual rebut the presumption of an at-will power to discharge . . . ."); Lowe v. Sorenson Research Co., 779 P.2d 668, 670 (Utah 1989) (summary judgment improper when trier of fact might find employer failed to follow handbook termination procedures); Gilmore v. Salt Lake Area Community Action Program, 775 P.2d 940, 941 (Utah Ct. App. 1989), *cert. denied*, 787 P.2d 33 (Utah 1990) (question of fact whether employee who received policy manual after beginning employment had implied contract).

166. 818 P.2d 997 (Utah 1991).

III. DISCLAIMERS AS A LIMITATION ON RECOVERY UNDER IMPLIED-IN-FACT CONTRACTS

A disclaimer is an express statement, usually contained in employee handbooks or on job applications, that an employee cannot rely on the employer's statements or conduct as a contractual commitment.[167] Courts[168] and commentators[169] have encouraged employers to use disclaimers to avoid wrongful termination claims and to preserve at-will employment. Employers have done so with mixed success.[170]

This section examines the split of authority on the enforceability of disclaimers in employee handbooks. Some courts, including the Utah Supreme Court in *Johnson*, have held that a disclaimer prohibits an employee from relying on an employee handbook for implied-in-fact contract provisions.[171] Other courts have refused to give such broad deference to disclaimers.[172] Instead, these courts have closely examined the employment relationship to determine if the disclaimer is consistent with other

---

167. *See* Henry H. Perritt, Jr., *The Future of Wrongful Dismissal Claims: Where Does Employer Self Interest Lie?*, 58 CINN. L. REV. 397, 417 (1989).

168. *See, e.g.*, Toussaint v. Blue Cross & Blue Shield, 292 N.W.2d 880, 891 n.24 (Mich. 1980) (employer can enter express at-will contract with employee); Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1271, *modified*, 499 A.2d 515 (N.J. 1985) (To avoid wrongful termination liability, employers should include statement that there are no promises in handbook, and regardless of what the handbook says an employer is free to change wages and working conditions at will.); Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1049-50 (Utah 1989) (Durham & Stewart, JJ.) (employer can overcome implied terms by demonstrating they were "expressly disavowed").

169. *See, e.g.*, Chagares, *supra* note 7, at 377 (disclaimer is useful tool in employment context); Kurt H. Decker, *Handbooks and Employment Policies as Express or Implied Guarantees of Employment—Employer Beware!*, 5 J.L. & COM. 207, 223 (1985)("To minimize liability, an employer should consider adding a disclaimer to its handbook."); Lenard, *supra* note 7, at 549 ("employer can disavow any of the apparent promises in such employee literature by publishing a prominent disclaimer with the material"); Perritt, *supra* note 167, at 417 (disclaimer "expressly reserving the employer's right to terminate with or without cause, and precluding informal agreements to the contrary, negates any possibility of an implied agreement to dismiss only on certain grounds").

170. *Compare* Novosel v. Sears, Roebuck & Co., 495 F. Supp. 344, 346 (E.D. Mich. 1980) (Sears' disclaimer precluded employees from forming legitimate expectations that they would be terminated only for cause) *with* Tiranno v. Sears, Roebuck & Co., 472 N.Y.S.2d 49, 50 (App. Div. 1984) (Sears' disclaimer did not preclude jury from finding just cause required for termination).

171. *See infra* notes 174-94 and accompanying text (discussing *Johnson* and cases similarly decided).

172. *See infra* notes 195-224 (discussing cases refusing to uphold disclaimers).

handbook provisions as well as the employer's de facto policies.[173] These two approaches are analyzed below.

A. Disclaimer Upheld: Johnson v. Morton Thiokol

In *Johnson v. Morton Thiokol, Inc.*,[174] the Utah Supreme Court, following the lead of other jurisdictions,[175] held that a disclaimer in an employee handbook preserves the employer's right to discharge for any reason. In *Johnson*, plaintiff Billy Johnson ("Johnson") sued his former employer Morton Thiokol ("Thiokol") for breaching an implied contract requiring that Thiokol terminate him only for cause. Thiokol employed Johnson as a process inspector.[176] One of his duties involved verifying the correct placement and seal of O-rings used in Thiokol's space shuttle rocket motors. On the day of the incident that lead to his termination, Johnson verified that he had completed the procedure, but failed to notice that certain hoses had been improperly connected. After the motor was seriously damaged during a testing operation,[177] Thiokol terminated Johnson in accordance with the

---

173. *See infra* notes 264-72 and accompanying text (discussing effect of employers' de facto policies).

174. 818 P.2d 997 (Utah 1991).

175. In the following cases, the employer invoked a disclaimer as a defense to an implied contract claim and the court upheld the disclaimer: Doe v. First Nat'l Bank, 865 F.2d 864, 873 (7th Cir. 1989); Dell v. Montgomery Ward & Co., 811 F.2d 970, 972-74 (6th Cir. 1987); McCluskey v. Unicare Health Facility, Inc., 484 So. 2d 398, 400 (Ala. 1986); Bennet v. Evanston Hosp., 540 N.E.2d 979, 980-81 (Ill. App. Ct. 1989); Moore v. Illinois Bell Tel., 508 N.E.2d 519, 521 (Ill. App. Ct.), *appeal denied*, 515 N.E.2d 112 (Ill. 1987); Nork v. Fetter Printing Co., 738 S.W.2d 824, 827 (Ky. Ct. App. 1987); Castiglione v. Johns Hopkins Hosp., 517 A.2d 786, 793-94 (Md. Ct. Spec. App.), *cert. denied*, 523 A.2d 1013 (Md. 1986); Bailey v. Perkins Restaurants, Inc., 398 N.W.2d 120, 121-23 (N.D. 1986); Messerly v. Asamera Minerals (U.S.), Inc., 780 P.2d 1327, 1330 (Wash. Ct. App. 1989).

The courts in the following cases suggested that a disclaimer would negate the contractual effect of a handbook, although a disclaimer was not at issue: Fletcher v. Wesley Medical Ctr., 585 F. Supp. 1260, 1264 (D. Kan. 1984); Hoffman-La Roche, Inc. v. Campbell 512 So. 2d 725, 734 (Ala. 1987); Pine River State Bank v. Mettille, 333 N.W.2d 622, 627 (Minn. 1983); Small v. Springs Indus., Inc., 357 S.E.2d 452, 455 (S.C. 1987).

176. The facts are taken from the court's opinion in *Johnson*, 818 P.2d at 998-99.

177. The damage led to a twenty-day delay in test firing the rocket motor. It also prompted a NASA investigation and adverse publicity, both nationally and locally. *Id.* at 999.

procedures set forth in its employee handbook.[178] Johnson then initiated the grievance procedures outlined in the handbook. Thiokol followed the prescribed procedures but denied Johnson relief. Johnson then sued for wrongful termination.

Johnson did not have an express employment contract with Thiokol. Thiokol's handbook, however, contained several pages describing Thiokol's disciplinary, appraisal, and grievance procedures.[179] The handbook was prefaced by the following disclaimer:

> This book is provided for general guidance only. The policies and procedures expressed in this book, as well as those in any other personnel materials which may be issued from time to time, do not create a binding contract or any other obligation or liability on the company. Your employment is for no set period and may be terminated without notice and at will at any time by you or the company. The company reserves the right to change these policies and procedures at any time for any reason.[180]

The trial court granted Thiokol's motion for summary judgment on the grounds that there was not an implied-in-fact contract between Johnson and Thiokol. On appeal, the supreme court reiterated that the terms of an employee handbook can operate as an implied-in-fact contract and establish the terms of

---

178. *Id.*

179. *Id.* at 998. The handbook stated, among other things,

> [i]t is the policy of Morton Thiokol, Inc., to establish reasonable rules of employment conduct and to ensure compliance with these rules through a program consistent with the best interests of the Company and its employees.
>
> . . . .
>
> PROMOTIONS AND PAY INCREASES
>
> Like most people, you are probably looking forward to increased responsibility and promotion, and Morton Thiokol recognizes this as a natural and worthy desire.
>
> Performance evaluations provide a guide for improvement of your work and a basis for your promotions and merit increases in pay.

Id. at 998-99 n.2. The text of the handbook also contained "several pages prescribing Thiokol's policy concerning employee disciplinary, appraisal, and grievance procedures." *Id.* at 998 (footnote omitted). Moreover, the handbook outlined grievance procedures that allowed terminated employees to challenge their discharge. *Id.* at 999.

180. *Id.* at 1003.

an employment relationship.[181] According to the court, the handbook constitutes the employer's offer for a unilateral contract.[182] The employee accepts this offer by continuing to work for the employer.[183] Consequently, the employer's handbook must meet the requirements of an offer for a unilateral contract.[184] In order to bind an employer, the handbook terms must evince "a manifestation of the employer's intent that is communicated to the employee and sufficiently definite to operate as a contract provision."[185] Moreover, an employee must reasonably believe that this manifestation of the employer's intent is an offer for something other than at-will employment.[186]

The court rejected Johnson's implied-in-fact contract claim because given the disclaimer, "the only reasonable conclusion an employee or a juror could reach concerning Thiokol's intention is that Thiokol intended to retain the right to discharge for any reason."[187] Therefore, in spite of the termination procedures outlined in the handbook, Thiokol could terminate Johnson at any time, for any reason.[188] Thus, *Johnson* holds that a disclaimer is an absolute bar to implied-in-fact contract terms based on employee handbooks.[189]

---

181. *Id.* at 1000 (citing *Berube*, 771 P.2d at 1044).
182. *Id.* at 1002.
183. *Id.*
184. *Id.*
185. *Id.* (footnotes omitted).
186. *Id.*
187. *Id.* at 1003.
188. *Id.*
189. *Id.*; *see also* Doe v. First Nat'l Bank, 865 F.2d 864, 872 (7th Cir. 1989) (employee cannot rely on handbook that contains disclaimer); Dell v. Montgomery Ward & Co., 811 F.2d 970, 973-74 (6th Cir. 1987) (same). The *Johnson* court agreed, however, that a subsequent express or implied agreement could modify the at-will employment, even if expressly disclaimed. *Johnson*, 818 P.2d at 1004 n.29. "Although an implied provision cannot contradict an express contract term . . . an express contract can be modified by a subsequent implied contract." *Id.* (citations omitted); *see also* Heslop v. Bank of Utah, 194 Utah Adv. Rep. 20, 20, 24 (Sept. 4, 1992) (express disclaimer in employment application modified by employers subsequent oral agreement); *infra* notes 264-72 and accompanying text (discussing effect of employer's de facto policies).

In Moore v. Illinois Bell Telephone Co., 508 N.E.2d 519 (Ill. App. Ct.), *appeal denied*, 515 N.E.2d 112 (Ill. 1987), discharged employees sued their former employer for breaching an implied employment contract. *Id.* at 519-20. The Illinois Court of Appeals held that a disclaimer in the handbook prevented reasonable reliance on the part of the employee and therefore reversed a jury verdict in the employee's favor. *Id.* at 521. According to the court, "it would not be reasonable for an employee to believe that [the employer] had made an offer in the [handbook], and so no enforceable contractual rights

Justice Stewart, concurring in *Johnson*, stated that the majority placed too much emphasis on the disclaimer.[190] Justice Stewart believed that the issue was not whether the disclaimer was a controlling manifestation of the employer's intent; rather, the issue was whether the disciplinary procedures outlined in the handbook modified the disclaimer.[191] According to Justice Stewart, because an employer's policies and procedures concerning discharge are sufficient evidence to overcome the presumption of at-will employment, the policies and procedures also should be sufficient to overcome express assertions that employment is at will.[192] Thus, disclaimers are effective only when they do not contradict the employer's discharge policies and procedures.[193] Justice Stewart agreed that summary judgment was properly granted, however, because it was undisputed that Thiokol had complied with the termination and review procedures set forth in its handbook.[194]

B. Cases Refusing to Uphold Disclaimers

A number of cases have refused to uphold disclaimers in circumstances similar to *Johnson*. These decisions have relied on a variety of legal principles to support their outcomes. Common to these decisions, however, is the fact that detailed procedures for discipline and termination accompanied the disclaimer. These decisions recognize that the employers have fostered reasonable expectations of job security, while improperly reserving the right to terminate at will.[195]

Some courts have held that disclaimers are vague and ambiguous, and therefore unenforceable. For example, in *Jones v. Central Peninsula General Hospital*,[196] the Alaska Supreme

---

were created by [it]." *Id.*; *see also* Castiglione v. Johns Hopkins Hosp., 517 A.2d 786, 793-94 (Md. Ct. Spec. App. 1986), *cert. denied*, 523 A.2d 1013 (Md. 1987) (justifiable reliance precluded where contractual intent disclaimed); *Doe*, 865 F.2d at 872 (same).

190. *Johnson*, 818 P.2d at 1005-06 (Stewart & Durham, JJ., concurring).

191. *Id.* at 1004 (Stewart & Durham, JJ., concurring).

192. *Id.* at 1005 (Stewart & Durham, JJ., concurring).

193. *Id.* at 1006 (Stewart & Durham, JJ., concurring).

194. *Id.* (Stewart & Durham, JJ., concurring).

195. *See, e.g.*, Jimenez v. Colorado Interstate Gas Co., 690 F. Supp. 977, 980 (D. Wyo. 1988) (strong equitable policy against allowing employers to mislead employees with policy statements while reserving right to terminate at will).

196. 779 P.2d 783 (Alaska 1989).

Court considered an employee handbook that purported to disclaim any contractual liability.[197] The provisions following the disclaimer provided, among other things, that the termination of nonprobationary employees would be "for cause" only.[198] The employer argued, however, that all of its employees were terminable at will in light of the disclaimer itself.[199] Yet, the court did not believe the disclaimer was clear and unambiguous: a "one-sentence disclaimer, followed by 85 pages of detailed text . . . does not unambiguously and conspicuously inform the employee that the manual is not part of the employee's contract of employment."[200] The manual's provisions created an impression of job security. The disclaimer, therefore, did not prevent the employee from using the manual to modify her at-will employment status.[201]

Closely related to the ambiguity principle is the idea that disclaimers are inherently inconsistent with handbook provisions guaranteeing job security. For example, in *Alexander v. Phillips Oil Co.*,[202] the Wyoming Supreme Court held that a handbook disclaimer[203] stating that employees could be terminated with or without cause was inconsistent with provisions that listed infractions warranting discharge.[204] The "overall tenor" of the handbook implied that cause was necessary for termination.[205]

---

197. *Id.* at 787. The disclaimer at issue stated: "The purpose of this manual is to provide information to all Society employees. It is not a contract of employment nor is it incorporated in any contract of employment between the Society and any employee." *Id.*

198. *Id.* at 785.

199. *Id.* at 787.

200. *Id.* at 788.

201. *Id.*; *see also* McDonald v. Mobil Coal Producing, Inc., 820 P.2d 986, 991 (Wyo. 1991) (despite disclaimer, employer's intent was ambiguous).

202. 707 P.2d 1385 (Wyo. 1985).

203. The disclaimer stated, in part:

> [The handbook] is intended to be used only as a general guide for benefits and policies as they may exist from time to time. The company retains the right to improve, discontinue, modify and/or interpret all or part of such policies and the contents of this booklet at any time without the prior knowledge or consent of the employee.
>
> . . . .
>
> . . . . [A]ny employee service can be terminated, with or without cause, at any time, at the option of either the company or the employee.

Id. at 1387.

204. *Id.* at 1388.

205. *Id.*

Thus, when considered as a whole, the handbook was sufficient evidence to support a modification of at-will employment.[206]

Other courts have looked at the employee's expectations. In *Zaccardi v. Zale Corp.*,[207] the United States Court of Appeals for the Tenth Circuit, applying Utah law, reversed the trial court's grant of summary judgment that was based on a disclaimer[208] in a handbook.[209] The court stated that a "disclaimer does not automatically negate a document's contractual status and must be read by reference to the parties' 'norms of conduct and expectations founded upon them.'"[210] Similarly, *Aiello v. United Air Lines, Inc.*,[211] involved an employee handbook with an introductory statement disclaiming[212] contractual liability.[213] The court noted that extensive regulations in handbooks could lead employees to believe that they are entitled to certain rights[214]—rights that the employer should not be able to revoke simply by pointing to a disclaimer.[215] Thus, rather than treating disclaimers as the controlling manifestation of employers' intent, these courts consider the reasonableness of the disclaimer in light of other handbook provisions.

Still other courts have given almost no rationale for refusing to uphold disclaimers. In *Kari v. General Motors Corp.*,[216] for example, the Michigan court of appeals held that a handbook

---

206. *Id.* at 1389.

207. 856 F.2d 1473 (10th Cir. 1988).

208. The disclaimer stated:

> While every effort has been made to create a manual which is specific as possible the policies and procedures set forth are a guideline and are not intended to cover and cannot cover every contingency, circumstance or situation. *This manual is not intended and shall not be interpreted to be a formal legal contract, binding on the company.*

Id. at 1476 n.5.

209. *Id.* at 1476-77.

210. *Id.* (quoting Hillis v. Meister, 483 P.2d 1314, 1317 (N.M. 1971)).

211. 818 F.2d 1196 (5th Cir. 1987).

212. The *Aiello* disclaimer was an introductory statement to the employer regulations and policies manual that stated: The regulations "are not intended to be, and do not constitute, a contractual arrangement or agreement between the company and its employees of any kind, . . . all employment is 'at-will.'" *Id.* at 1198.

213. *Id.*

214. *Id.* at 1200 n.2 (citing Cleary v. American Airlines, 168 Cal. Rptr. 722 (Dist. Ct. App. 1980)).

215. *Id.*

216. 261 N.W.2d 222 (Mich. Ct. App. 1977), *rev'd*, 282 N.W.2d 925 (Mich. 1978).

disclaimer[217] was sufficient to avoid liability for a terminated employee's severance pay.[218] The Michigan Supreme Court reversed in a one-sentence opinion, stating simply "the trial court erred in granting a summary judgment without hearing testimony in this case."[219]

The above cases represent judicial response to disclaimers that purportedly bestow absolute employer discretion.[220] Despite their varying rationale, the theme of these cases is that employers cannot point to a disclaimer after they have enjoyed the benefits of a loyal and cooperative workforce created by guaranteeing job security. In other words, employers cannot have their cake and eat it too.[221] Rather than announce broad rules of decision, rife with the possibility of future misapplication,[222] these courts closely scrutinize the employment relationship for inequity.[223] This

---

217. The disclaimer stated:

> The contents of this handbook are presented as a matter of information only. While General Motors believes wholeheartedly in the plans, policies and procedures described here, they are not conditions of employment. General Motors reserves the right to modify, revoke, suspend, terminate, or change any or all such plans, policies, or procedures, in whole or in part, at any time, with or without notice. The language used in this handbook is not intended to create, nor is it to be construed to constitute, a contract between General Motors and any one or all of its employees.

Id. at 223.

218. *Id.* at 224.

219. Kari v. General Motors Corp., 282 N.W.2d 925, 925 (Mich. 1978); *see also* Aiello v. United Air Lines, Inc., 818 F.2d 1196, 1200 (5th Cir. 1987) ("It is a truism in the law that such disavowals are not controlling."). Other courts have looked closely at whether the employer brought the disclaimer to the terminated employee's personal attention. *See* Morris v. Coleman Co., 738 P.2d 841, 849 (Kan. 1987).

220. *See* Cynthia W. Scherb, Note, *The Use of Disclaimers to Avoid Employer Liability Under Employee Handbook Provisions*, 12 J. CORP. L. 105, 118 (1986); *see also* Wagenseller v. Scottsdale Memorial Hosp., 710 P.2d 1025, 1037-38 (Ariz. 1985) ("[H]aving announced a policy, the employer may not treat it as illusory.").

221. *See* Scherb, *supra* note 220, at 119.

222. *See* Peterson v. Browning, 832 P.2d 1280, 1292 (Utah 1992) (Zimmerman, J., & Hall, C.J., dissenting) ("We write our decisions in the abstract, using a supposed set of facts in our attempt to draw a careful line between the permissible and the impermissible. Once they leave our chambers, however, our rules operate in the real world of inexactitude and compromise, where judges and juries are not always consistent . . . .").

223. *See generally Wagenseller*, 710 P.2d at 1038 (intent to create non-at-will relationship is discerned from totality of the parties' statements and actions); Loffa v. Intel Corp., 738 P.2d 1146, 1150 (Ariz. Ct. App. 1987) (whether handbook modified at-will employment is question of fact to be determined from totality of parties' statements and employment relationship); Wilkerson v. Wells Fargo Bank, 261 Cal. Rptr. 185, 189 (Ct. App. 1989) (factors apart from express terms may be used to ascertain existence and

approach leaves courts with the flexibility necessary to dispense with cases summarily when warranted.[224] More importantly, it allows the courts to grant redress to employees when necessary.

IV. A BALANCED APPROACH TO EMPLOYEE HANDBOOK DISCLAIMERS

In deciding whether to enforce a handbook disclaimer, a court should carefully balance the often competing interests of the employer, the employee, and the public.[225] Employers have a legitimate interest in operating their businesses without undue governmental intrusion.[226] Unnecessary limitations of an employer's ability to discharge employees impinge on the employer's ability to manage effectively its operations. On the other hand, employees have an interest in knowing that they can rely on an employer's assurance of job security.[227] Employees, after all, have the greatest stake in the employment relationship. Finally, the public has an interest in discouraging groundless lawsuits by disgruntled employees.[228] These lawsuits burden both the judicial system and the economy.

Johnson and similar cases protect employers' interests in managing their operations and the public's interest in discouraging litigation. However, they cut deeply into employees' interest in being able to rely on employers' representations of job security. This Note proposes that courts employ a balanced analysis of

---

content of agreement, including personnel policies or practices, employee's longevity of service, employer's actions reflecting assurance of continued employment, and industry practices); Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1265, *modified*, 499 A.2d 515 (N.J. 1985) (to determine handbook's meaning, court must consider context in which it was disseminated and environment surrounding its continued existence); Tiranno v. Sears, Roebuck & Co., 472 N.Y.S.2d 49, 50 (App. Div. 1984) (jury must consider totality of circumstances to find if just cause required to terminate employee).

224. *See, e.g.*, Longley v. Blue Cross & Blue Shield, 356 N.W.2d 20, 22-23 (Mich. Ct. App. 1984) (upholding summary judgment for employer where employee knew she was at-will, even though employment conditions indicated termination for just cause); Ferraro v. Koelsch, 368 N.W.2d 666, 674-76 (Wis. 1985) (employee's implied contract requiring termination for cause not breached when employer had cause for termination).

225. *See Woolley*, 491 A.2d at 1261 (citing Pierce v. Ortho Pharmaceutical Corp., 417 A.2d 505, 509 (N.J. 1985)).

226. *See id.*

227. *See id.*

228. *See id.*

handbook disclaimers; distinguishing valid statements of employer policy from invalid, boilerplate disclaimers.

In developing an alternative approach, this section first critically examines the *Johnson* decision and the practical effect of its holding, arguing that rather than furthering the policies underlying the implied contract exception to the at-will rule, the court's approach instead encourages bad faith. Next, this section outlines an alternative approach to handbook disclaimers that more accurately reflects the employment relationship and protects employees' reasonable expectations.

A. *The Impact of* Johnson v. Morton Thiokol

In *Johnson*, a majority of the Utah Supreme Court expressed a willingness to elevate boilerplate disclaimers above the reasonable expectations of employees. As a result, Utah employers are now encouraged to promulgate handbooks with detailed job security provisions while knowing that these provisions are unenforceable if accompanied by a disclaimer. This result is contrary to the spirit and policy behind the implied-in-fact contract exception recognized by the Utah Supreme Court in *Berube*.[229]

The employment-at-will doctrine was adopted to deal with a rapidly industrializing society, and most jurisdictions followed the rule without adequately considering countervailing interests.[230] Among the several factors that led to the erosion of the at-will rule, the most important was the realization that it often was harshly applied.[231] In reexamining the at-will rule many courts, including the Utah Supreme Court, rejected the rule's theoretical underpinnings and recognized it merely as a presumption, rather than a statement of substantive law.[232] The primary reason for judicially limiting the at-will rule is that the rule allows employers to make illusory promises to employees.[233] As one court noted,

---

229. 771 P.2d 1033, 1040-42 (Utah 1987) (Durham & Stewart, JJ.); *id.* at 1050 (Howe, J., & Hall, C.J., concurring); *id.* at 1052-53 (Zimmerman, J., concurring).

230. *Id.* at 1041 (Durham & Stewart, JJ.); Gary E. Murg & Clifford Scharman, *Employment At Will: Do the Exceptions Overwhelm the Rule?*, 23 B.C. L. REV. 329, 335 (1982).

231. *See* Murg & Scharman, *supra* note 230, at 340.

232. *Berube*, 771 P.2d at 1040-42 (Durham & Stewart, JJ.); *id.* at 1050 (Zimmerman, J., concurring).

233. Small v. Springs Indus., Inc., 357 S.E.2d 452, 455 (S.C. 1987).

"[i]t is patently unjust to allow an employer to couch a handbook . . . in mandatory terms and then allow him to ignore these very policies as a 'gratuitous, nonbinding statement of general policy' whenever it works to his disadvantage."[234] It is, of course, equally unjust to allow employers to make grandiose promises of job security and then avoid liability by the simple expedient of a boilerplate disclaimer.

According to *Berube*, the guiding principle in construing employment contracts is giving effect to the parties' intentions.[235] Although *Johnson* purported to adhere to this principle, the court decided, as a matter of law, that handbook disclaimers are the controlling manifestation of an employer's intent.[236] Although this rationale logically applies to employers who hire at-will employees and notify them of their status through disclaimers, it is distinguishable from employers who guarantee job security while simultaneously disclaiming their responsibility to grant it. In the former situation, the employer is merely advising its employees of the legal presumption.[237] In the latter, the employer hopes to gain the benefits of a loyal, motivated, and non-unionized workforce, and at the same time insulate itself from liability for wrongfully terminating employees. Enforcing disclaimers in the latter situation encourages deception.[238]

B. An Alternative Analysis

A balanced approach requires that a court distinguish enforceable statements of policy from unenforceable, boilerplate disclaimers. This section sets out an analytical framework for making such a distinction. First, this section considers the policy considerations behind handbook disclaimers. Second, this section distinguishes disclaimers that satisfy these policy considerations and therefore should be enforced from those that fail this test and

---

234. *Id.*

235. *See Berube*, 771 P.2d at 1044 (Durham & Stewart, JJ.); *see also Johnson*, 818 P.2d at 1003 (court must read handbook as a whole, harmonizing all provisions).

236. *Johnson*, 818 P.2d at 1003. Given the language of the disclaimer, "the only reasonable conclusion an employee or a juror could reach concerning Thiokol's intention is that Thiokol intended to retain the right to discharge for any reason." *Id.*

237. *Id.* at 1000 (indefinite-term employee is presumed employee-at-will and thus terminable for any reason).

238. Finkin, *supra* note 16, at 750.

therefore should not. In essence, this distinction involves comparing the handbook as a whole with the disclaimer. If the disclaimer is inconsistent with other provisions in the handbook, a court should declare the disclaimer unenforceable. If the disclaimer is consistent, it simply reiterates the employer's policy and should be enforced. Finally, this section considers the related issue of the impact of an employer's de facto policies on the enforceability of handbook disclaimers. These policies, treated as subsequent assurances of job security, also should override express handbook disclaimers.

1. Policy Considerations

Nonunionized, private-sector employers have a legitimate interest in maintaining broad discretion with respect to terminating employees.[239] Some commentators argue that because unionized employees enjoy significant job security guarantees, courts should completely abolish the at-will rule and thereby grant nonunionized employees the same protections.[240] There is a critical difference, however, between job security granted in the union context and job security created through exceptions to the at-will rule. In the union context, job security is arrived at thorough collective bargaining, often obtained at the expense of strikes costly to employees. In contrast, exceptions to the at-will rule are externally imposed at no cost to individual employees.[241] Although this distinction does not go unchallenged, the refusal to abolish completely the at-will rule indicates that society perceives a difference between unionized and at-will employees.[242]

---

239. Hodges v. Gibson Prod. Co., 811 P.2d 151, 165 (Utah 1991) (Stewart & Durham, JJ.).

240. *See* Peck, *supra* note 14, at 728-31. Professor Peck argues that the growing recognition of exceptions to the at-will rule is intended to provide the same level of protection to at-will employees that is enjoyed by unionized and public employees. *Id.* at 720. The legal doctrines that the courts have used to fashion exceptions to the at-will rule, however, are ill-suited to accomplish this objective. *Id.* Peck refers to the manipulation of these legal doctrines as "doctrinal camouflage." *Id.* The real explanation for this doctrinal camouflage, Peck argues, is the incompatibility of the at-will rule with modern society. *Id.*

241. *See* Lopatka, *supra* note 51, at 5.

242. *See generally* Murg & Scharman, *supra* note 230, at 331 (abolishing at-will rule "would fundamentally alter the traditional employer-employee relationship, and would impose enormous burdens on employers").

Because of this difference, employers use job security guarantees to avoid unionization and ensure an orderly, loyal, and cooperative workforce.[243] The implied contract exception to the at-will rule prevents an employer from making illusory promises by treating such guarantees as enforceable contracts. Overly broad recognition of disclaimers, however, seriously undermines the purpose of the implied contract exception.

Employers often argue that failure to give absolute deference to disclaimers will have a chilling effect on employers' willingness to disseminate handbooks.[244] According to this argument, the uncertainty and possibility of judicial second-guessing that would accompany less-than-absolute deference to disclaimers essentially would dictate that even employers who wished to issue handbooks refrain from doing so.[245] This argument, however, ignores the substantial benefits that handbooks provide to employers, even in an at-will setting. Handbooks are useful methods for communicating nonpromissory policies to employees.[246] They can communicate information such as appropriate and inappropriate behavior, the structure of a company's operations, and the company's business philosophy.[247] Perhaps most importantly, employers can also use handbooks to ensure that employees understand their employment is at will. Handbooks in an at-will employment setting, therefore, actually benefit both employer and employee.[248] Thus, even if disclaimers were not accorded absolute deference, employers would still have a strong incentive to draft and disseminate handbooks.

2. *Distinguishing Enforceable from Unenforceable Disclaimers*

An employer is not required to establish personnel policies or practices. When an employer does not establish policies, the law will presume that its employees are at will.[249] However, when

---

243. *See* Scherb, *supra* note 220, at 118.

244. *See* Brief of Respondent at 17, Johnson v. Morton Thiokol, Inc., 818 P.2d 997 (Utah 1991) (No. 890315).

245. *See id.*

246. *See* Chagares, *supra* note 7, at 376-78.

247. *See id.* at 377.

248. *See id.*

249. *See* Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1041 (Utah 1989) (Durham & Stewart, JJ.).

"an employer chooses to establish such policies and practices and makes them known to its employees, the employment relationship is presumably enhanced."[250] This enhancement is the product of employee reliance on representations of job security. When an employer chooses to take advantage of this workforce tranquility, the employer should be willing to pay the cost; namely, the employer should be bound by the policies it has promulgated.[251] Courts should not condone offering inducements and benefits that the employer can withdraw simply by pointing to a disclaimer.[252]

Whenever an employer invokes a disclaimer to avoid implied contract liability, courts should closely scrutinize the entire handbook, as well as the text of the disclaimer. If an employer has made promises of job security in the handbook, those promises should be read in light of the disclaimer. In doing so, the court should give a reasonable construction to the handbook as a whole.[253] Viewing these facially incompatible provisions together, the issue should not be whether the disclaimer is a controlling manifestation of the employer's intent;[254] rather, the issue should be whether the right to discharge at will is limited by the job security guarantees.[255] If the job security guarantees are limitations, the terms of the implied contract, and whether the employer breached those terms, are questions to be resolved by a trier of fact. If, on the other hand, examination of the handbook and disclaimer reveals that the employer clearly and unequivocally

---

250. Toussaint v. Blue Cross & Blue Shield, 292 N.W.2d 880, 892 (Mich. 1980).

251. *See* Finkin, *supra* note 16, at 751-52 ("When employers act to foster a belief in job security and fair treatment to avoid high turnover (and unionization), they pay a price in being bound by the assurances they have given and by the rules they have fashioned toward those ends.").

252. Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1266, *modified*, 499 A.2d 515 (N.J. 1985). One commentator sums it up: "[E]mployers cannot expect to oversell employment opportunities, publish slick employee handbooks with reassuring job security statements, or make grandiose claims of fairness and enlightenment in an undiscriminating effort to keep unions out and land the best personnel, without substantial risk of wrongful discharge liability." Lopatka, *supra* note 51, at 27.

253. *See Johnson*, 818 P.2d at 1003; *cf.* Sears v. Riemersma, 655 P.2d 1105, 1107-08 (Utah 1982) ("The primary rule in interpreting a contract is to determine what the parties intended by looking at the entire contract and all of its parts in relation to each other, giving an objective and reasonable construction to the contract as a whole."); Baker v. Western Sur. Co., 757 P.2d 878, 881 (Utah Ct. App. 1988) (discussing importance of intent with respect to interpretation of contract).

254. *Johnson*, 818 P.2d at 1003.

255. *Id.* at 1004 n.1 (Stewart & Durham, JJ., concurring).

established an at-will relationship, courts should enforce that arrangement. Presumably, an employer willing openly to announce this policy will attract the type of employees willing to work under it.

An employer wishing to use disclaimers to notify its employees of their at-will status should consider the substantive content as well as the appearance of the disclaimer. The disclaimer should be clear, unambiguous, and explicit.[256] One commentator has suggested that such clarity involves indicating three items: "(1) that the employment relationship is terminable at the will of either party, (2) that it is terminable with or without cause, and (3) that it is terminable without prior notice."[257] However, a disclaimer is only useful to advise employees of their at-will status if the handbook unambiguously informs them of their status. Consequently, the most important requirement for an enforceable disclaimer is that it be consistent with other handbook provisions.[258]

For example, an employer may state on the first page of its handbook that the information contained in the handbook is for informational purposes only and is not a contract of employment, but in subsequent provisions of the handbook represent that employees will be terminated for cause only. In this situation, the employer has not unambiguously conveyed to its employees that they are employed at will.[259] Likewise, a handbook disclaimer followed by a list of causes for which termination will occur does not unambiguously convey at-will employment.[260]

---

256. *See* Leikvold v. Valley View Community Hosp., 688 P.2d 170, 174 (Ariz. 1984); Castiglione v. Johns Hopkins Hosp., 517 A.2d 786, 793 (Md. Ct. Spec. App. 1986); Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1258, *modified*, 499 A.2d 515 (N.J. 1985); Thompson v. St. Regis Paper Co., 685 P.2d 1081, 1088 (Wash. 1984).

257. Chagares, *supra* note 7, at 384 (footnote omitted).

258. *See* Clare Tully, Note, *Challenging the Employment-At-Will Doctrine Through Modern Contract Theory*, 16 U. MICH. J.L. REF. 449, 463 (1983). The disclaimer must also be consistent with the employer's de facto policies. *See infra* notes 264-72 and accompanying text (discussing effect of employer's de facto policies on employment-at-will relationship).

259. *See supra* notes 196-201 and accompanying text (discussing Jones v. Central Peninsula Gen. Hosp., 779 P.2d 783 (Alaska 1989)).

260. *See supra* notes 202-06 and accompanying text (discussing Alexander v. Phillips Oil Co., 707 P.2d 1385 (Wyo. 1985)).

For this reason, employers should remove any language in the handbook that uses or suggests job security.[261] Employers must delete any inaccurate or ambiguous descriptions of "permanent" employment, detailed grievance or disciplinary procedures to be taken prior to termination, and lists of exclusive reasons for discharge to ensure that the handbook does not implicitly guarantee job security.[262]

Of course, an employer may avoid either extreme and seek instead to provide limited assurances of job security. Thiokol, for example, guaranteed that it would follow certain disciplinary and grievance procedures prior to terminating employees.[263] The employer must recognize, however, that failure to abide by these procedural guarantees subjects the employer to liability.

3. Effect of De Facto Policies

The effect of an employer's unwritten or de facto policies and procedures is closely related to the enforceability of disclaimers. These de facto policies should be treated as subsequent guarantees of job security, and thus should also negate the effect of an express disclaimer. As the *Johnson* court recognized, an express at-will relationship can be modified by a subsequent implied contract.[264] Courts should be willing to construe employers' unwritten policies and procedures as giving rise to such an implied contract.[265]

Moreover, when considering an at-will relationship, a court should not limit itself to the language of the handbook and accompanying disclaimer. Instead, a court should consider the circumstances surrounding the employment relationship.[266] Even when the disclaimer is unambiguous and prominently placed, it should be only one factor in a court's inquiry and should be viewed in light of workplace reality.[267] Courts need to be willing to look

---

261. Decker, *supra* note 169, at 224.

262. *Id.*; Chagares, *supra* note 7, at 392.

263. *See supra* notes 179-80 and accompanying text (discussing Thiokol's disciplinary and grievance procedures).

264. Johnson v. Morton Thiokol, Inc., 818 P.2d 997, 1004 n.29 (Utah 1991); *see also* Heslop v. Bank of Utah, 194 Utah Adv. Rep. 20, 20, 24 (Sept. 4, 1992) (express disclaimer in employment application modified by subsequent oral agreement).

265. *See Johnson*, 818 P.2d at 1004 (Stewart & Durham, JJ., concurring).

266. *See* Lenard, *supra* note 7, at 562.

267. *See id. But cf. Johnson*, 818 P.2d at 1003 (clear and conspicuous disclaimer

to factors such as employers' actions that reflect assurances of job security, employers' unwritten personnel policies and practices, employees' performance and longevity of service, and industry practices.[268] These de facto policies may combine to overcome an express statement in a handbook that employment is at will.

The Utah Supreme Court has demonstrated a willingness to give effect to such subsequent assurances of job security. In *Heslop v. Bank of Utah*,[269] the court recognized that even though the employment application signed by the discharged employee contained a disclaimer,[270] the employer's subsequent promise that it would terminate the employee only for good cause bound the employer.[271] Thus, if the employer's de facto policies indicate that all terminations will be for cause only, the at-will relationship is modified and the employer is bound by an implied contract.[272]

V. CONCLUSION

In broadly upholding handbook disclaimers, the *Johnson* decision fails to address the legitimate expectations of Utah employees. The unfortunate result is that employers are encouraged to promulgate handbooks replete with assurances of job security, knowing that these provisions are wholly unenforceable. Rather than sacrifice employee interests, courts confronted with handbook disclaimers should distinguish valid statements of policy embodied in disclaimers from invalid, boilerplate disclaimers—giving effect to the former while leaving the validity of the latter to be weighed by a trier of fact. It is only through adopting such a policy that employers will be encouraged to fashion

---

prevents handbook from being considered implied-in-fact contract as a matter of law).

268. *See* Wilkerson v. Wells Fargo Bank, 261 Cal. Rptr. 185, 189 (Ct. App. 1989).

269. 194 Utah Adv. Rep. 20 (Sept. 4, 1992).

270. The disclaimer in *Heslop* stated:

> I agree that my employment will depend upon usefulness to the bank, in its sole discretion; the bank reserving the right to release me without notice, its obligation ending with the payment of salary through the last day I work.

Id. at 20.

271. *Id.* at 24 (holding statute of frauds did not bar subsequent oral contract).

272. *Id.* at 1004 & n.29. The employee's continued employment is sufficient consideration to support the modification. Berube v. Fashion Centre, Ltd., 771 P.2d 1033, 1045 (Utah 1989) (Durham & Stewart, JJ.); Caldwell v. Ford, Bacon & Davis Utah, Inc., 777 P.2d 483, 485-86 (Utah 1989).

reasonable policies with which both employer and employee can live.

TODD M. SHAUGHNESSY