PRINCIPLES OF BIOMEDICAL ETHICS

Fourth Edition

TOM L. BEAUCHAMP
JAMES F. CHILDRESS

Principles of Biomedical Ethics

174.2
B372
Ed.14

Principles of Biomedical Ethics

FOURTH EDITION

Tom L. Beauchamp
James F. Childress

UNIVERSIDAD RAFAEL LANDIVAR
BIBLIOTECA
Guatemala, C. A.

New York *Oxford*
OXFORD UNIVERSITY PRESS
1994

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland Madrid

and associated companies in
Berlin Ibadan

Copyright © 1979, 1983, 1989, 1994 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.,
200 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
Beauchamp, Tom L.
Principles of biomedical ethics -
Tom L. Beauchamp, James F. Childress.—4th ed.
p. cm. Includes bibliographical references and index.
ISBN 0-19-508536-1.—ISBN 0-19-508537-X (pbk.)
1. Medical ethics. I. Childress, James F. II. Title.
[DNLM: 1. Ethics, Medical.]
R724.B36 1994
174'.2—dc20 93-24390

9 8 7 6 5 4 3 2

Printed in the United States of America
on acid-free paper

To
Georgia, Ruth, and Don

I can no other answer make but thanks,
And thanks, and ever thanks.

Twelfth Night

#32.95
3-14-95

Preface to the Fourth Edition

Biomedical ethics was a young field when the first edition of this book went to press in late 1977. Immense changes occurred in the field's literature between the first edition and the present, fourth edition. Although major changes have appeared in all editions after the first, this edition includes more significant changes than any other.

Readers will notice particularly far-reaching changes in Chapters 1, 2, and 8, which contain our main views in ethical theory. In Chapter 1, entirely new sections have been added on Method, Justification, and Truth; Specifying and Balancing Principles; and The Place of Principles. In Chapter 2, new sections have been developed on Liberal Individualism, Communitarianism, The Ethics of Care, Casuistry, and Principle-Based, Common-Morality Theories. In Chapter 8, new sections appear on Virtues in Professional Roles, Four Focal Virtues, and Moral Excellence. In other chapters, major sections have been added on The Justification of Assistance in Dying, The Value and Quality of Life, and Rationing through Priorities in the Health Care Budget. Many substantive changes also appear in sections that are not new to this edition. For example, in the Fair Opportunity section of Chapter 6, we have added a subsection on Distributing Health Care on the Basis of Gender and Race.

Despite these changes, the book retains its previous chapter structure and characteristic perspectives on major issues. Most of the chapter and sectional headings have also been retained from the third edition.

Because of the concentration on theory in the first two chapters, some readers may prefer to read the chapters devoted to various principles (Chapters 3–6), rules (Chapter 7), and virtues (Chapter 8) before exploring our approaches to theory and method.

We have also altered the way we present cases in this edition. More complete versions of the relevant cases are now found in the text, rather than the appendix. Thus, although the appendix of cases is now smaller, more discussion of cases is provided throughout the text.

We have received many helpful suggestions for improvements in the previ-

ous edition from students, colleagues, health professionals, and teachers who use the book. We owe special thanks in this edition to David DeGrazia and Henry Richardson for their assessments of new sections in Chapters 1 and 2, and to Ruth Faden for her reading of Chapters 5 and 6. We are also grateful to John Hasnas and Madison Powers for their criticisms of the new section on casuistry.

Several articles critical of aspects of our book have appeared in the literature of biomedical ethics in recent years. Although we do not always concur with our critics, we are especially grateful to our friends John Arras, Dan Clouser, Bernie Gert, and Ron Green for some very probing and often penetrating suggestions.

Special acknowledgment is due to a talented research staff that has assisted us in the collecting of materials. Catherine Marshall, Brian Tauscher, Liz Emmett, and Felicia Cohn provided invaluable assistance and advice in this process, as well as in the revision of drafts of chapters. They repeatedly forced us to make our sections more pertinent and readable. Emily Wilson, Andrew Dodge, and Brian Marshall helped with page proofs and the index. In addition, our university offices provided superb assistance by faithfully preparing draft after draft. We are especially indebted for this assistance to Moheba Hanif and Diana McKenzie.

Some parts of chapters were presented at the seminars of the Kennedy Institute of Ethics. Many arguments were changed as a result of the comments and responses of our critics on these occasions. We also acknowledge with due appreciation the support provided by the Kennedy Institute's library and information retrieval systems, which kept us in touch with new literature and reduced the burdens of library research. In particular, we thank Mary Coutts for her faithful and precise searching of data bases.

Washington, D.C. T.L.B.
Charlottesville, Virginia J.F.C.
November 1993

Contents

1. Morality and Moral Justification, 3

Morality and Ethical Theory, 4
Moral Dilemmas, 11
Method, Justification, and Truth, 13
Specifying and Balancing Principles, 28
The Place of Principles, 37
Conclusion, 40

2. Types of Ethical Theory, 44

Criteria for Theory Construction, 45
Utilitarianism: Consequence-Based Theory, 47
Kantianism: Obligation-Based Theory, 56
Character Ethics: Virtue-Based Theory, 62
Liberal Individualism: Rights-Based Theory, 69
Communitarianism: Community-Based Theory, 77
Ethics of Care: Relationship-Based Accounts, 85
Casuistry: Case-Based Reasoning, 92
Principle-Based, Common-Morality Theories 100
Convergence Across Theories, 109
Conclusion, 111

3. Respect for Autonomy, 120

The Concept of Autonomy, 120
Competence and Autonomous Choice, 132
The Meaning and Justification of Informed Consent, 142
Disclosure, 146
Understanding, 157
Voluntariness, 163
A Framework of Standards for Surrogate Decisionmaking, 170
Conclusion, 181

4. Nonmaleficence, 189

The Concept of Nonmaleficence, 190
Traditional Distinctions and Rules Governing Nontreatment, 196
Optional Treatments and Obligatory Treatments, 211
Killing and Letting Die, 219
The Justification of Assistance in Dying, 235
Decisionmaking for Incompetent Patients, 241
Conclusion, 249

5. Beneficence, 259

The Concept of Beneficence, 260
Obligatory and Ideal Beneficence, 261
Paternalism: Conflicts between Beneficence and Autonomy, 271
Balancing Benefits, Costs, and Risks, 291
The Value and Quality of Life, 305
Conclusion, 317

6. Justice, 326

The Concept of Justice, 327
Theories of Justice, 334
Fair Opportunity, 341
The Right to a Decent Minimum of Health Care, 348
The Allocation of Health Care Resources, 361
Rationing through Priorities in the Health Care Budget, 366
Rationing Scarce Treatments to Patients, 378
Conclusion, 386

7. Professional–Patient Relationships, 395

Veracity, 395
Privacy, 406
Confidentiality, 418
Fidelity, 429
The Dual Roles of Physician and Investigator, 441
Conclusion, 452

8. Virtues and Ideals in Professional Life, 462

Virtues in Professional Roles, 463
Four Focal Virtues, 466
Conscientiousness, 474
Moral Ideals, 483
Moral Excellence, 490
Conclusion, 502

Appendix: Cases in Biomedical Ethics, 509

Index, 527

Principles of Biomedical Ethics

1

Morality and Moral Justification

Medical ethics enjoyed a remarkable degree of continuity from the days of Hippocrates until its long-standing traditions began to be supplanted, or at least supplemented, around the middle of the twentieth century. Scientific, technological, and social developments during that time produced rapid changes in the biological sciences and in health care. These developments challenged many prevalent conceptions of the moral obligations of health professionals and society in meeting the needs of the sick and injured. The objective of this book is to provide a framework for moral judgment and decisionmaking in the wake of these developments.

Although major writings in ancient, medieval, and modern health care contain a rich storehouse of reflection on the relationship between the professional and the patient, this history is often disappointing from the perspective of contemporary biomedical ethics. It shows how inadequately, and with what measure of insularity, problems of truthfulness, privacy, justice, communal responsibility, and the like were framed in past centuries. To avoid a similar inadequacy, we begin with a study of ethics that may appear to be distant from both history and contemporary problems in the biological sciences, medicine, nursing, and other modes of health care. Our objective is to show how ethical theory can illuminate problems in health care and can help overcome some limitations of past formulations of ethical responsibility. However, it is unreasonable to expect any theory to overcome all the limitations of time and place and reach a universally acceptable perspective.

Morality and Ethical Theory

Ethics is a generic term for various ways of understanding and examining the moral life. Some approaches to ethics are normative (that is, they present standards of right or good action), others are descriptive (that is, they report what people believe and how they act), and still others analyze the concepts and methods of ethics.

Approaches to Ethics

Normative ethics. Inquiry that attempts to answer the question "Which general norms for the guidance and evaluation of conduct are worthy of moral acceptance and for what reasons?" is *general normative ethics*. The study of these norms has typically occurred while developing an ethical theory, although the concept of "normative ethics" is a distinctly twentieth-century notion not invoked in pre-twentieth-century theories. Ideally, such a theory satisfies a set of criteria for theories that is developed in Chapter 2, thereby making it a comprehensive inquiry into moral concepts, principles, reasoning, and the like. However, numerous practical questions would remain unanswered even if a fully satisfactory general ethical theory were available, for reasons explained at the end of this chapter.

The attempt to work out the implications of general theories for specific forms of conduct and moral judgment will be called *practical ethics* here, although it is often misleadingly called *applied ethics*. The term *practical* refers to the use of ethical theory and methods of analysis to examine moral problems, practices, and policies in several areas, including the professions and public policy. Often no straightforward movement from theory or principles to particular judgments is possible in these contexts, although general reasons, principles, and even ideals can play some role in evaluating conduct and establishing policies. Theory and principles are typically invoked only to help develop action-guides, which are also further shaped by paradigm cases of appropriate behavior, empirical data, and the like, together with reflection on how to put these influential sources into the most coherent whole.

Although general normative ethics is distinguished from practical ethics by being more general and by a weaker connection to concerns about practical affairs, no sharp distinction should be drawn between the two. In this book we are mainly concerned with interpreting principles and developing general moral action-guides for use in the biomedical fields, but we will also examine moral motives and character. (*Biomedicine* is a shorthand expression for the biological sciences, medicine, and health care.)

Nonnormative ethics. In addition to normative ethics, there are two broad types of nonnormative ethics. First, *descriptive ethics* is the factual investigation of moral behavior and beliefs. It uses standard scientific techniques to study how people reason and act. For example, anthropologists, sociologists, psychologists, and historians determine which moral norms and attitudes are expressed in professional practice, in codes, and in public policies. They study different beliefs and practices regarding surrogate decisionmaking, treatment of the dying, the nature of consent obtained from patients, and the like.

Second, *metaethics* involves analysis of the language, concepts, and methods of reasoning in ethics. For example, it addresses the meanings of such ethical terms as *right, obligation, virtue, principle, justification, sympathy, morality,* and *responsibility*. It also includes study of moral epistemology (the theory of moral knowledge) and the logic and patterns of moral reasoning and justification. Metaethical questions for analysis include whether social morality is objective or subjective, relative or nonrelative, and rational or emotive.

Descriptive ethics and metaethics are grouped together as nonnormative because their objective is to establish what factually or conceptually is the case, not what ethically ought to be the case. Often in this book we cite descriptive ethics—for example, by presenting what professional codes require. However, the underlying question is usually whether the described prescriptions of such codes are defensible, which is a normative issue. We also frequently engage in metaethics. Chapter 1 is largely an exercise in metaethics, whereas Chapter 2 is devoted to general normative ethics. However, such distinctions should be used with caution.[1] Metaethics frequently takes a sharp turn toward the normative, as our discussion of the justification of moral standards later in this chapter indicates. Likewise, normative ethics relies heavily on metaethics. Just as no sharp distinction should be drawn between practical ethics and normative ethics, so no clear line should be drawn to distinguish normative ethics and metaethics.

The Common Morality

In addressing the question "What is morality?" we may be tempted to answer that morality is a theory about right and wrong. However, the words *ethics* and *morality* should not be confined to theoretical contexts. *Ethical theory* and *moral philosophy* are the appropriate terms to refer to philosophical reflection on morality's nature and function. The purpose of theory is to enhance clarity, systematic order, and precision of argument in our thinking about morality. The term *morality* refers to social conventions about right and wrong human conduct that are so widely shared that they form a stable (although usually incomplete) communal consensus, whereas *ethics* is a general term referring to both morality and ethical theory. (The terms *ethical* and *moral* are here construed as identical in meaning.) *Ethical theory, moral philosophy,* and *philosophical ethics,* then, are reserved for philosophical theories, including reflec-

tion on the common morality. Similarly, *moral theology, theological ethics,* and *religious ethics* are reserved for reflection on morality in light of theological convictions in particular religious traditions.

In its broadest and most familiar sense, the *common morality* comprises socially approved norms of human conduct. For example, it recognizes many legitimate and illegitimate forms of conduct that we capture by using the language of "human rights." The common morality is a social institution with a code of learnable norms. Like languages and political constitutions, the common morality exists before we are instructed in its relevant rules and regulations. As we develop beyond infancy, we learn moral rules along with other social rules, such as laws. Later in life, we learn to distinguish general social rules held in common by members of society from particular social rules fashioned for and binding on the members of special groups, such as the members of a profession.

The common morality is not faultless or complete in its recommendations, but we will later argue that it forms the right starting point for ethical theory.

Codes of Professional Ethics

Influential reflection on problems of biomedical ethics within the health care professions has evolved through formal codes of medical and nursing ethics, codes of research ethics, and reports by government-sponsored commissions. *Particular* codes written for groups such as physicians, nurses, and psychologists are sometimes defended by appeal to general norms such as not harming others (nonmaleficence) and respecting autonomy and privacy, even if these were not explicitly considered in the drafting of the codes. Codes can be validly criticized or defended by appeal to general norms, as can many public policies and regulations that have been formulated to guide professionals' conduct.

Before we assess professional codes, the nature of professions needs brief discussion. According to Talcott Parsons, a profession is "a cluster of occupational roles, that is, roles in which the incumbents perform certain functions valued in the society in general, and by these activities, typically earn a living at a full-time job."[2] Under this definition circus performers, exterminators, and waitresses are professionals; prostitutes probably are not (because their function is not "valued in the society in general"), despite prostitution's reputation as "the world's oldest profession." However, it is not surprising to hear prostitution characterized as a profession, inasmuch as the word *profession* has come, in common use, to mean almost any occupation in which a person earns a living.

We need a more restricted meaning for the term *profession*, as used in *professional ethics*. Professionals are usually identified by their commitment to provide important services to clients or consumers and by their specialized

training. Professions maintain self-regulating organizations that control entry into occupational roles by formally certifying that candidates have acquired the necessary knowledge and skills. The concept of a medical professional is closely tied to a background of distinctive education and skills that patients typically lack and that morally must be used to benefit patients. In learned professions such as medicine the background knowledge of the professional derives from closely supervised training, and the professional is one who provides a service to others. However, not all professions are either learned or service-oriented.

Health care professions typically specify and enforce obligations, thereby seeking to ensure that persons who enter into relationships with their members will find them competent and trustworthy. The obligations that professions attempt to enforce are role obligations that are correlative to the rights of other persons. Problems of professional ethics usually arise from conflicts of values, sometimes conflicts within the profession and sometimes conflicts of professional commitments with commitments of persons outside the profession. A professional code represents an articulated statement of the role morality of the members of the profession, and in this way professional standards are distinguished from standards imposed by external bodies such as governments (although their norms sometimes overlap and agree).

Codes also often specify rules of etiquette and responsibilities to other members of the profession. For example, one historically significant code of the American Medical Association instructed physicians not to criticize a fellow physician previously in charge of a case and urged all physicians to offer professional courtesy.[3] These codes tend to foster and reinforce member-identification with and institutional conformity to the prevailing values of the profession. These professional codes are beneficial if they effectively incorporate defensible moral norms. Unfortunately, some professional codes oversimplify moral requirements or claim more completeness and authority than they are entitled to claim. As a consequence, professionals may mistakenly suppose that they satisfy all moral requirements if they obediently follow the rules of the code, just as many people believe that they discharge their moral obligations when they meet all relevant legal requirements.

A pertinent question concerns whether the codes specific to areas of science, medicine, and health care are comprehensive, coherent, and plausible in their moral norms. Many medical codes develop the implications of some general principles, such as "Do no harm," and of some rules, such as rules of medical confidentiality. But only a few have anything to say about the implications of other principles and rules, such as veracity, respect for autonomy, and justice, which have been the subjects of intense contemporary discussion. Some of these neglected principles and rules have recently been incorporated into statements of patients' rights that invoke, for example, the principle of respect for

autonomy and rules of veracity.[4] These statements of proper professional conduct do not derive from professional codes and differ from many codes by focusing on the *rights* of those receiving health services rather than on the *obligations* of health professionals. Unfortunately, such statements are also usually incomplete and lack an argued defense.

Reasons other than incompleteness and lack of stated justification also support skepticism about the adequacy of professional codes in health care. From the time of Hippocrates, physicians have generated codes without scrutiny or acceptance by patients and the public. These codes have rarely appealed to more general ethical standards or to a source of moral authority beyond the traditions and judgments of physicians. In some cases, the special rules in codes for professionals seem to conflict with and even override more general moral norms. The pursuit of professional norms in these circumstances may do more to protect the profession's interests than to introduce an impartial and comprehensive moral viewpoint. Other rules have traditionally been expressed in abstract formulations that dispense vague moral advice open to competing interpretations.

Related reservations about codes of medical ethics were poignantly expressed in 1972 by psychiatrist Jay Katz. Originally inspired by his outrage over the fate of Holocaust victims, Katz became convinced that only a persistent educational effort beyond traditional codes could provide meaningful guidance in research involving human subjects:

> As I became increasingly involved in the world of law, I learned much that was new to me from my colleagues and students about such complex issues as the right to self-determination and privacy and the extent of the authority of governmental, professional, and other institutions to intrude into private life. . . . These issues . . . had rarely been discussed in my medical education. Instead it had been all too uncritically assumed that they could be resolved by fidelity to such undefined principles as *primum non nocere* or to visionary codes of ethics.[5]

Government Guidelines and Public Policy

Additional moral direction for health professionals and scientists is sometimes provided through the public policy process, which includes specific regulations and guidelines promulgated by government agencies. Public policies, such as those that fund health care for the indigent and those that protect subjects of biomedical research, usually incorporate moral considerations. Moral analysis is part of good policy formation, not merely a method for evaluating already formed policy.

Rights of patients and research subjects provide instructive examples. The U.S. government has promulgated several legally binding regulations intended to provide morally adequate protections for research subjects. In 1974, Congress created a national commission to recommend research guidelines to the secretary of the Department of Health, Education and Welfare (now the Department of Health and Human Services) that would become federal regulations unless the secretary publicly justified not implementing them. In 1980, a president's commission was assembled to further examine these issues in research and other issues, such as access to health care and decisionmaking in clinical-patient relations.[6] Finally, in December 1991, the Patient Self-Determination Act (PSDA) went into effect.[7] Congress passed this law as the first federal legislation to ensure that health care institutions inform patients about their rights under state law and about institutional policies to accept or refuse medical treatment and to formulate advance directives.

Several U.S. federal agencies and courts regularly use ethical premises in the development of their health policies, rules, decisions, or analyses. These agencies include the Centers for Disease Control (CDC), the National Institutes of Health (NIH), the Office of Technology Assessment (OTA), and the U.S. Supreme Court. On the state level, formal ethical analysis also often plays a prominent role in policy formation in bioethics. Examples include the widely examined work of the New York Task Force on Life and the Law and of the New Jersey Bioethics Commission. Such commission reports and government acts, along with other government policies pertaining to biomedicine, raise vital questions explored later in this book about the proper relation between government and professional groups in formulating standards of practice.

These questions are also raised by the prominent role of the courts in the United States in developing case law that sets standards for science, medicine, and health care. These legal decisions have been significant resources for ethical reflection about moral responsibilities and public policy. Examples include decisions about informed consent (see Chapter 3) and terminating life-sustaining treatment (see Chapter 4). In the latter, the line of court decisions since the Karen Ann Quinlan case in the mid-1970s constitutes a nascent moral tradition of reflection that has been influenced by extralegal writings on topics such as whether artificial nutrition and hydration should be viewed as a medical treatment subject to the same standards of decisionmaking as other forms of treatment.

The term *public policy* is used in this text to refer to a set of normative, enforceable guidelines that have been accepted by an official public body, such as an agency of government or a legislature, to govern a particular area of conduct. The policies of corporations, hospitals, trade groups, and professional societies sometimes have a deep impact on public policy; but these policies are private rather than public (although these bodies are frequently regulated by public policies). A far closer connection exists between law and public policy, inasmuch as all laws constitute public policies; but not all public policies are, in the conventional sense, laws. In contrast to laws, public policies need not

be explicitly formulated or codified. For example, an official who decides not to fund a program that has no prior history of funding formulates a public policy. Decisions not to act as well as decisions to act can constitute public policies.

Policy formation and criticism involve more complex forms of judgment than merely invoking ethical principles and rules.[8] The ethics of public policy must proceed from impure and unsettled cases, in which there are profound social disagreements, uncertainties, different interpretations of history, and imperfect procedures to resolve the disagreements. Obviously, no body of abstract principles and rules can dictate policy, because it cannot contain enough specific information or provide direct and discerning guidance. The specification and implementation of moral principles and rules must take account of problems of feasibility, efficiency, cultural pluralism, political procedures, uncertainty about risk, noncompliance by patients, and the like. Principles and rules provide background moral considerations for policy evaluation, but a policy must also be shaped by empirical data and by special information available in relevant fields of medicine, economics, law, psychology, and so on. In this process, moral principles and various uses of empirical data are often closely connected. For example, techniques such as risk–benefit analysis are not purely empirical and value-free, because they involve moral evaluation and need constraint by principles of justice.

Finally, when using moral principles or rules to formulate or criticize public policies, we cannot move with assurance from a judgment that *act x* is morally right (or wrong) to a judgment that a *law* or *policy y* is morally right (or wrong) because it mandates or encourages (or prohibits) act *x*. The judgment that an act is morally wrong, then, does not necessarily lead to the judgment that the government should prohibit it or refuse to allocate funds to support it. For example, one can consistently argue that sterilization or abortion is morally wrong without holding that the law should prohibit it or deny public funds to those who otherwise could not afford the procedure.

Similarly, the judgment that an act is morally acceptable does not imply that the law should permit it. For example, the thesis that active euthanasia is morally justified if patients face uncontrollable pain and suffering and request death is consistent with the thesis that the government should legally prohibit active euthanasia because it would not be possible to control abuses if it were legalized. We are not here defending particular moral judgments about the justifiability of such acts. We are maintaining only that the connections between moral action-guides and judgments about policy or law or legal enforcement are complicated and that a judgment about the morality of acts does not entail a particular judgment about law and policy. Factors such as the symbolic value of law, the costs of a program and its enforcement, and the demands of competing programs must also be considered.

Moral Dilemmas

Facing and reasoning through dilemmas to conclusions and choices is a familiar feature of the human condition. Consider a particular case (Case 1 in the appendix). Some years ago the judges on the California Supreme Court had to reach a decision about a possible violation of medical confidentiality. A man killed a woman after confiding to a therapist his intention to commit the act. The therapist attempted unsuccessfully to have the man committed but, in accordance with his duty of medical confidentiality to the patient, did not communicate the threat to the woman when the commitment attempt failed. The majority opinion of the court held that "When a therapist determines, or pursuant to the standards of his profession should determine, that his patient presents a serious danger of violence to another, he incurs an obligation to use reasonable care to protect the intended victim against such danger." Accordingly, this obligation extends to notifying the police and directly warning the intended victim. The justices in the majority opinion argue that therapists generally ought to observe the rule of medical confidentiality but that this rule must yield in this case to the "public interest in safety from violent assault." Although they recognize that rules of professional ethics have substantial public value, they hold that matters of greater importance, such as protecting others against violent assault, can override the rules.

In a minority opinion, a justice disagrees with this analysis and argues that doctors violate patients' rights if they fail to observe standard rules of confidentiality. If it were common practice to break these rules, he reasons, the fiduciary nature of the relationship between physicians and patients would begin to erode. The mentally ill would refrain from seeking aid or divulging critical information, because of the loss of trust that is essential for effective treatment. As a result, violent assaults would more than likely increase. This case presents a straightforward moral dilemma (as well as a legal dilemma), because both judges cite good and relevant reasons to support their conflicting judgments.

Moral dilemmas occur in at least two forms.[9] (1) Some evidence indicates that act *x* is morally right, and some evidence indicates that act *x* is morally wrong, but the evidence on both sides is inconclusive. Abortion, for example, is sometimes said to be a terrible dilemma for women who see the evidence in this way. (2) An agent believes that, on moral grounds, he or she both ought and ought not to perform act *x*. In a moral dilemma with this form, an agent is obligated by one or more moral norms to do *x* and obligated by one or more moral norms to do *y*, but the agent is precluded in the circumstances from doing both. The reasons behind alternatives *x* and *y* are good and weighty, and neither set of reasons is obviously dominant. If one acts on either set of reasons, one's actions will be morally acceptable in some respects but morally

unacceptable in others. Some have viewed the intentional cessation of lifesaving therapies in the case of persistent vegetative state patients, such as Karen Ann Quinlan and Nancy Cruzan, as dilemmatic in this second way.

Dilemmas can be created by conflicting moral principles and rules, as popular literature, novels, and films about difficult choices often illustrate. For example, an impoverished person steals to save a family from starvation, a mother kills one of her children to save a second child, or a person lies to protect a family secret. The only way to comply with one obligation in such situations is by contravening another obligation. No matter which course is elected, some obligation must be set aside. However, it is both awkward and misleading to say that we are, in the circumstances, obligated to perform both actions. We should discharge the obligation that in the circumstances overrides what we would have been firmly obligated to perform were it not for the conflict.

Conflicts between moral requirements and self-interest sometimes produce a *practical* dilemma that is not a *moral* dilemma. If moral reasons compete with nonmoral reasons, difficult questions about priority can still be posed even though moral dilemmas are not present. Numerous examples appear in the work of anthropologist William R. Bascom, who collected hundreds of "African Dilemma Tales" transmitted for decades or centuries in African tribal societies. One traditional dilemma posed by the Hausa tribe of Nigeria is called *cure for impotence:*

> A friend gave a man a magical armlet that cured his impotence. Later he saw his mother, who had been lost in a slave raid, in a gang of prisoners. He begged his friend to use his magic to release her. The friend agreed on one condition—that the armlet be returned. What shall his choice be?[10]

Hard choice? Perhaps, but not a hard *moral* choice. The obligation to the mother is moral in character, whereas retaining the armlet is a matter of self-interest. (We are assuming that no moral obligation exists to a sexual partner; in some circumstances such an obligation could produce a moral dilemma.)

Some moral philosophers and theologians have argued that many types of *practical* dilemmas exist, but never genuine *moral* dilemmas. They do not deny that agents experience moral perplexity, moral conflict, and moral disagreement in difficult cases, but they insist that if there were moral dilemmas two moral *oughts* would conflict, so that one could not perform an action that one ought to perform without forgoing another action one also ought to perform. The belief that one *cannot* do what one *ought* to do seems to these writers a confusion about the nature of moral language and obligation. Some major figures in the history of ethics have defended this conclusion, both because they accept one supreme moral value as overriding all other conflicting values (moral and nonmoral) and because they regard it as incoherent to allow contradictory oughts. The only *ought*, they maintain, is the *ought* generated by the supreme value.[11]

By contrast, we will maintain that various moral principles can and do conflict in the moral life. On some occasions the conflict produces a moral dilemma with no supreme principle to determine an overriding *ought*. Nonetheless, there are ways of reasoning about what should be done. In some cases the dilemma can be resolved, but in other cases the dilemma only becomes more difficult and remains unresolved after careful reflection.

Method, Justification, and Truth

A person of good will typically has no difficulty in making moral judgments about telling the truth and avoiding untenable conflicts of interest. Our routine moral judgments are reached through a mix of appeals to guidelines, models, parables, and the like. These moral beacons usually suffice, because we are not asked to deliberate about or justify either our judgments or the standards that underlie them. However, when we experience moral perplexity or uncertainty, we are led to moral reasoning about what morality recommends or requires of us and to deliberation about what we should do. From there we often find a need for moral justification.

Deliberation is primarily problem solving in which individuals or groups struggle to develop and assess their beliefs in order to reach a decision. As John Dewey once observed, deliberation begins with "an imaginative rehearsal of various courses of action."[12] As we deliberate, we usually consider which among the possible courses of action is morally justified, i.e. which has the strongest moral reasons behind it. The reasons we finally accept express the conditions under which we believe some course of action is morally justified.

But what, more precisely, is justification in ethics, and by what method of reasoning do we achieve it? *Justification* has several meanings in English, some specific to disciplines such as theology and law. In its customary sense, to justify is to show to be right, to vindicate, to furnish adequate grounds for, to warrant, and the like. In some theological traditions, justification is the condition under which a person is freed from sin and made righteous. In law, justification is a showing in court that one has a sufficient reason for one's claim or for what one has been called to answer. In ethical discourse, the legal sense provides the closest analogue. The objective is to establish one's case by presenting sufficient grounds for belief and action. To demonstrate that one is justified in a moral belief requires that one make explicit one's underlying warrants for the belief. A mere listing of beliefs will not suffice, because an *alleged* reason often provides no support for an intended conclusion. Not all reasons are good reasons, and not all good reasons are sufficient for justification. There is, then, a need to distinguish a reason's *relevance* to a moral judgment from its final *adequacy* for that judgment, and also to distinguish an *attempted* justification from a *successful* justification.

For example, the presence of dangerous toxic chemicals in a work environment was widely offered by chemical companies in the United States as a legally and morally sound reason to exclude women of childbearing age from a hazardous workplace, but in 1991 the United States Supreme Court overturned these policies as discriminatory.[13] The dangers to health and life presented by hazardous chemicals constitute a *good* reason for removing employees from the workplace, but this reason may not be a *sufficient* reason for a ban exclusively directed at women. No matter which position is defended, we expect proponents to give us a further account of why their reasons amount to both good and sufficient reasons.

Three Models of Justification

Several models of justification are operative in ethical theory. We will evaluate three that are both instructive and influential. The first approaches justification from a top-down perspective that emphasizes general norms and ethical theory as the proper basis for reaching correct moral judgments. The second approaches justification from a bottom-up perspective that emphasizes moral tradition, experience, and judgment as the bases of both general norms and theory. The third refuses to assign priority to either a top-down or a bottom-up strategy. We will discuss each approach and then defend a version of the third that incorporates important parts of the other two.

Deductivism: the covering-precept model. The covering-precept model—or, as it is now widely called, *deductivism*—holds that justified moral judgments are deduced from a preexisting theoretical structure of normative precepts that cover the judgment. This model is inspired by justification in disciplines such as mathematics, in which a claim is shown to follow logically (deductively) from a credible set of premises. This mode of argument is occasionally used in ethics, and deductivists promote it as the best model of justification. The idea is that justification occurs if and only if general principles and rules, together with the relevant facts of a situation, support an inference to the correct or justified judgment(s). This model is simple and conforms to the way virtually all persons first learn to think morally: Moral judgment is the application of a rule (principle, ideal, right, etc.) to a clear case falling under the rule. The deductive form is therefore sometimes said to be a top-down "application" of general precepts—a conception that motivated use of the term "applied ethics." The deductive form in the application of a rule is the following (here using what is *obligatory*, rather than what is *permitted* or *prohibited*, although the deductive model is the same for all three):

1. Every act of description *A* is obligatory.
2. Act *b* is of description *A*.

Therefore,

3. Act *b* is obligatory.

Using this schematic model, consider the following example:

1x. Every act in a patient's overall best interest is obligatory for the patient's doctor.[14]

2x. Act of resuscitation *b* is in this patient's overall best interest.

Therefore,

3x. Act of resuscitation *b* is obligatory for this patient's doctor.

Covering precepts, as in 1 and 1x, occur at various levels of generality. A particular judgment, belief, or hypothesis may be justified by bringing it under one or more moral rules, or the rules may be justified by bringing them under principles, or both rules and principles may be defended by a full ethical theory. As an example, consider a nurse who refuses to assist in an abortion procedure. The nurse might attempt to justify the act of refusal by the rule that it is wrong to kill an innocent human being intentionally. If pressed for further justification, the nurse may justify this moral rule by reference to a principle of the sanctity of human life. Finally, the particular judgment, the rule, and the principle might all be justified by an ethical theory that is only implicit and inchoate in the nurse's original judgment. From a pure deductivist perspective, the primary problem of practical ethics is the choice of an ethical theory to apply. Presumably each competitive theory commits a proponent to different norms and different solutions of problems.

This model, then, directs attention from the level of particular judgments to a covering level of generality (rules and principles that cover and justify particular judgments), and then on to the level of ethical theory (which covers and warrants rules and principles). This account can be diagrammed as follows (where each arrow describes a direction of justification, indicating that the less general assertion is justified by appeal to a covering, more general norm):

4. Ethical Theory

↑

3. Principles

↑

2. Rules

↑

1. Particular Judgments

This model functions smoothly whenever a judgment can be brought directly under a rule or a principle without intervening complexities such as appeals to several principles.

This model of justification does not capture how moral reasoning and justification proceed in complicated cases. Oversimplification appears in its linear conception, in which particular judgments and rules are derivative in direct descent from more general principles. Consider the apparently deductivist justification, "You must tell Mr. Sanford that he has cancer and will probably die soon, because rules of truthfulness are essential in order to properly respect the autonomy of patients." The above model suggests that "You should not lie to Mr. Sanford" descends in its moral content directly from the covering principle "You should respect autonomy" and the covering rule "You shouldn't lie to patients." While much in the moral life does conform loosely to this linear-dependence conception, much also does not. For several reasons, actual moral reasoning and justification exhibit greater complexity.

First, general action-guides are reciprocally related to particular elements in experience. The relationship between general norms and the particulars of experience is bilateral, not unilateral; and, as Jerome Schneewind has argued, there is no "context-free order of dependence among moral propositions."[15] Whether a moral proposition is known by inference or is known independent of inference cannot be decided by either the content of the proposition or its degree of generality. Whether a general action-guide depends on particular experiences, or the converse, is a matter of what is known and inferred in specific contexts (so-called inferential support as a matter of epistemic context). Moral beliefs arise both by generalization from the particulars of experience and by making judgments in particular circumstances by appeal to general precepts. No essential order of inference or dependence fixes how we come to have moral knowledge.

Particular moral judgments also often involve specifying and balancing norms for concrete situations, not merely bringing a particular judgment under a covering rule or principle. The abstract rules and principles in moral theories are extensively indeterminate; that is, the act descriptions used to point to obligations lack sufficient determinative content for many practical judgments. In the process of specifying and balancing norms and in making particular judgments, we often must consider factual beliefs about the world, cultural expectations, judgments of likely outcome, precedents previously encountered, and the like to help fill out and give weight to rules, principles, and theories.

Moral judgments about the justifiability of abortion, for example, typically depend less on moral rules and principles than on beliefs about the nature and development of the fetus. Disputants agree on the general rule that it is wrong to kill innocent persons directly, but traditional customs, scientific theories, metaphysical convictions, conceptual accounts, and religious beliefs often underlie interpretations of situations. As a result, moral debate sometimes stems more from disagreement about the correct scientific, metaphysical, or religious description of the situation or from different conceptual accounts than from disagreement about the relevant moral action-guides.

The facts of a situation are often such that no general norm (principle or rule) clearly applies, even if we do not disagree about the rules or principles of morality. For example, destroying a nonviable human fetus is not a clear violation of rules against killing or murder, and the rule that one has a right to protect one's bodily integrity and property similarly does not clearly apply to the problem of abortion. Typically, the facts are complex, and several different moral norms can be brought to bear on the facts, but with inconclusive and even contradictory results. In the case of abortion, even if we have our facts straight, the choice of facts and the choice of rules will generate a judgment that is incompatible with another choice of facts and rules. Obtaining the right set of facts and bringing the right set of rules to bear on these facts are not reducible to either a deductive form of judgment or to the resources of a general ethical theory.

A further problem is that deductivism creates a potentially infinite regress of justification—a never-ending demand for final justification—because each level of appeal to a covering precept requires some further level to justify it. This problem can perhaps be handled by presenting a principle that is self-justifying or one that it is irrational not to hold. But if a deductivist demand for justification were to entail that every standard used to justify another standard is itself unjustified until brought under a different covering precept, then there cannot in principle be an adequate foundation for moral judgments or for the institution of morality.

Finally, the appeal to a level of theory in the covering-precept model suggests that only one correct normative theory exists, yet many distinct theories have been developed and ably defended, without a substantial consensus as to which system best meets the tests for a theory. To the surprise of many philosophers in the last twenty years, often little is lost in practical moral decisionmaking by dispensing with general moral theories. The rules and principles shared across these theories typically serve practical judgment more adequately (as starting points) than the theories. This paradox provides one major reason why covering-precept theories have recently diminished in influence. However, the model of covering precepts applied in moral judgments can be retained as long as it is complemented by other models that capture the greater complexity of moral reasoning and justification.

Inductivism: the individual-case model. Many people now believe, and we agree, that moral justification proceeds inductively (bottom-up) no less than deductively (top-down). Inductivism maintains that we must use existing social

agreements and practices as a starting point from which to generalize to norms such as principles and rules, and inductivists emphasize the role of particular and contextual judgments as a part of our evolving moral life. A society's moral views are not justified by an ahistorical examination of the logic of moral discourse or by some theory of rationality, but rather by an embedded moral tradition and a set of procedures that permit new developments. A static or morally conservative conception of morality does not follow from this account as long as the tradition presents methods and procedures for reflection on and development of the tradition. New experiences and innovations in the patterns of collective life lead to modifications in beliefs, and the institution of morality cannot be separated from a cultural matrix of beliefs that has grown up and been tested over time.

Inductivists argue that induction (reasoning from particular instances to general statements about the instances), including analogy (a species of induction in which similarities between acts or events support a hypothesis that the acts or events are similar in other respects as well), is central to deliberation and justification. They propose that certain kinds of cases and particular judgments about those cases can be relied upon as warrants for the acceptance of moral conclusions independently of either general norms or a historical tradition. As our experience and thinking develop, we then generalize beyond such judgments to create rules and principles that carry over to analogous contexts, but the inductivist believes that these rules and principles are *derivative* in the order of knowledge, not primary. New and revealing cases lead us to progressive refinements of our judgments and generalizations. Moral rules, then, are provisionally secure points in a cultural matrix of guidelines.

Consider an example from the explosion of interest since 1976 in surrogate decision making. A series of cases beginning with the aforementioned Quinlan case challenged medical ethics and the courts to develop virtually an entire new framework of substantive rules for responsible surrogate decisionmaking about life-sustaining treatments, as well as authority rules regarding who should make those decisions. This framework was created by working through cases analogically, and testing hypotheses against preexisting norms. In both ethics and law, a string of cases with some similar (and some dissimilar) features set the terms of the ethics of surrogate decisionmaking. Even when a principle or rule was not entirely novel in a proposed framework, its content was shaped by problems needing resolution in the cases. Gradually a loose consensus emerged in the courts and in ethics about a framework for such decisionmaking. It would falsify history to say that this framework was already available and was simply applied to new cases.

Some inductivists incorporate features of deductivism, understood as the claim that preexisting norms inform the process of analogy. This breadth makes inductivism more attractive, but not beyond criticism. There remains a certain

haziness about the role of particular experience and individual judgment. If the judgments of individuals are basic, do general norms have any critical power to correct biased judgments or biases that might be built into the norms generalized from particular experiences? What could give general rules authority over particular judgments?

During a discussion of this problem, Henry Sidgwick argues that if "the particular case can be satisfactorily settled by conscience [i.e., practical judgment] without reference to general rules," then "we shall have no *practical* need of any such general rules." And if we were able to "form general propositions by induction from these particular conscientious judgments, and arrange them systematically . . . any interest which such a system may have will be *purely speculative*," not practical.[16] Sidgwick's insight cuts deeply against pure inductivism, and he builds on the point by arguing that we use general norms to constrain and evaluate particular moral judgments. Theoretical interest in the general, then, is not entirely for the sake of a speculative typology of what we already know from particular cases. We are interested in general standards because we are interested in knowing what we ought to do.

A familiar problem haunts pure inductivism. We frequently criticize inadequate judgments or traditions by appeal to general standards such as human rights. What justification could inductivism provide for our use of these general standards if they stand outside the framework of experience and judgment being criticized? Inductivism also has little to say about the common circumstance in which conflicting judgments are reached in particular cases by equally well-informed and dispassionate moral agents. Inductivist theory, then, needs to be buttressed by an account of the proper role of rules and principles in adjudicating disputes and constraining particular judgments.

Before we look at the third of our three approaches to see if it can meet this condition, we need to summarize how far we have come. We have discussed two broad approaches to method and justification. When the limits of these approaches are acknowledged, we see no inconsistency between them and no reason to reject them outright. Inductivism rightly emphasizes that history and philosophy do not produce static systems of moral norms. We constantly engage in reasoned decisionmaking, moving from new experiences and problems to new or more refined action-guides. Deductivism rightly notes that once we have a fairly settled (although not necessarily final) body of general guidelines, moral judgments are often warranted by direct appeal to those general guidelines.

A confusing feature of contemporary moral theory is that false rivalries and misleading statements of method often result from pigeonholing theories too readily, and assuming that the proponents of one method exclude the other methods in their moral thinking. Here are two examples involving writers in biomedical ethics to whom we return later (see pp. 94–96 and 106–107). First,

Bernard Gert and Danner Clouser express their method as follows: "In formulating theory we start with particular moral judgments about which we are certain, and we abstract and formulate the relevant features of those cases to help us in turn to decide the unclear cases."[17] This statement is arresting because Gert and Clouser have gained a reputation (whether deserved or not) as full-fledged deductivists. Yet this statement of method looks more like inductivism.

Second, Albert R. Jonsen and Stephen Toulmin have gained a reputation (whether deserved or not) as leading proponents of an inductivism that emphasizes analogical reasoning, case-by-case, from already well–navigated moral territory—a method they call *casuistry*. Yet they describe "good casuistry" as an account that "applies general rules to particular cases with discernment."[18] Jonsen forthrightly supports a strategy that delineates general principles that are powerful enough to provide direction in new territory in which analogies seem insufficient or perilous. For example, he maintains that our cumulative experience and case reports are often insufficient to provide answers to emerging moral problems presented in new areas such as reproductive technologies and genome mapping.[19] Toulmin similarly acknowledges that principles have a central role to play in contexts in which the actors are strangers rather than intimates, as occurs increasingly in health care institutions.[20] In these reflections, their methodology looks more deductivist than inductivist, although final assessment depends on whether they view principles as entailing or directly supporting particular moral judgments.

An important lesson can be learned about the use of these labels. It is easy to mislabel and stereotype philosophical methods. Once the label is attached, the theory can be altogether dismissed through objections such as those we have offered. We again emphasize that we find much more that is acceptable than unacceptable in the work of authors who are committed to these broad approaches to justification, and we will draw from both. At the same time, we do not regard even both approaches together as adequate. We will spend the remainder of this chapter depicting the third, and our favored, account of justification.

Coherentism. "Coherentism," as the third approach may be called, is neither top-down nor bottom-up; it moves in both directions. John Rawls has used the term *reflective equilibrium* to refer to the goal of this form of justification, and we will adopt some central features in his analysis. Rawls views the acceptance of theory in ethics as properly beginning with our "considered judgments," the moral convictions in which we have the highest confidence and believe to have the lowest level of bias. Rawls's term *considered judgments* refers to "judgments in which our moral capacities are most likely to be displayed without distortion." Examples are judgments about the wrongness of racial discrimination, religious intolerance, and political representation. These considered judgments occur at all levels of generality in our moral thinking, "from those about particular situations and institutions through broad standards and first principles to formal and abstract conditions on moral conceptions."[21]

A considered judgment of long-standing prominence in medicine is the rule, "A physician must not exploit patients for the physician's own gain because the patient's interests come first." This rule implies that it is inappropriate to permit certain physician interests to coexist with a fundamental commitment to the patient. For example, entrepreneurial physicians with ownership or investment interests in laboratories and for-profit medical centers have an unwarranted conflict of interest if they create self-referral arrangements with centers in which they have a financial interest. This conduct in medicine is unprofessional, although it is acceptable practice in many areas of business. Businesspersons are not always constrained by a rule such as "the client's and customer's interest comes first." One difference between many medical and business codes of ethics springs from the above rule, a considered judgment, about the patient–physician relationship.

However, even the considered judgments that we accept "provisionally as fixed points" are, Rawls argues, "liable to revision." The goal of reflective equilibrium is to match, prune, and adjust considered judgments so that they coincide and are rendered coherent with the premises of theory.[22] That is, we start with paradigm judgments of moral rightness and wrongness, and then construct a more general theory that is consistent with these paradigm judgments (so that they are as coherent as possible); any loopholes are closed, as are all forms of incoherence that are detected. The resultant action-guides are then tested to see if they too yield incoherent results. If so, they are readjusted or given up, and the process is renewed, because we can never assume a completely stable equilibrium. The pruning and adjusting occur by reflection and dialectical adjustment, in view of the perpetual goal of reflective equilibrium. To refer again to the rule about putting the patient's interest first, we would seek in biomedical ethics to make this rule as coherent as possible with other considered judgments about clinical teaching responsibilities, responsibilities in the conduct of research involving human subjects, and responsibilities to patients' families.

As we argue in Chapter 7, it is difficult to bring these diverse commitments into coherence along with other moral rules of impartiality and fairness. The rule about putting the patient's interest first is an acceptable starting premise, but it is not categorical for all possible cases. We are left with a range of options about how we should and should not specify and balance the rule. As long as contingent conflicts occur under the recognized and legitimate principles and rules of a moral system or theory, some measure of incoherence is present. For example, to take a relatively uncomplicated example in the ethics

of organ transplantation, imagine that we are attracted to each of the following two moral considerations: (1) Distribute organs by expected number of years of survival, and (2) Distribute organs by time on the waiting list in order to give every candidate an equal opportunity. As they stand, these two distributive principles are not coherent, because use of either will undercut or eliminate the other. We can retain both (1) and (2) in a defensible theory of fair distribution, but to do so we will have to introduce limits on these principles as well as accounts of how to balance them. These limits and accounts will, in turn, have to be made coherent with other principles and rules, such as norms regarding discrimination against the elderly and the role of ability to pay in the allocation of expensive medical procedures.

This analysis suggests—rightly, we believe—that all moral systems present some level of indeterminateness and incoherence, revealing that they do not have the power to eliminate various contingent conflicts among principles and rules. So understood, coherence and reflective equilibrium are not achieved merely by an absence of inconsistencies in a system. Coherence is a matter of the further development and mutual support of norms.

So-called wide reflective equilibrium occurs when we evaluate the strengths and weaknesses of all plausible moral judgments, principles, and relevant background theories. That is, we incorporate as wide a variety of kinds of legitimate moral beliefs as possible, including hard test cases in experience.[23] We emphasize again the ideal (although not utopian) character of this procedure: No matter how wide the pool of beliefs, there is no reason to anticipate that the process of pruning, adjusting, and rendering coherent will either come to an end or be perfected. Virtually any set of theoretical generalizations achieved by reflective equilibrium will fall short of full coherence with considered judgments, and the only relevant model for moral theory is the *best approximation to full coherence*. We should assume that we are confronted with a never-ending search for defects of coherence, for counterexamples to our beliefs, and for novel situations.[24]

Ethical theories presumably can be made to cohere with considered judgments through this process of reflective equilibrium without incorporating controversial theoretical commitments about what is rational and irrational to accept. That is, many strands in the moral life can be reflectively considered, and brought into equilibrium, without introducing views about the rationality or irrationality of theories that deeply divide contemporary philosophers and theologians. From this perspective, moral thinking is analogous to hypotheses in science that are tested, modified, or rejected through experience and experimental thinking. Justification is neither purely deductivist (giving general action-guides preeminent status), nor purely inductivist (giving experience and analogy preeminent status). Many different considerations provide reciprocal support in the attempt to fit moral beliefs into a coherent unit. This is how we

test, revise, and further specify moral beliefs. This outlook is very different from deductivism, because it holds that ethical theories are never sufficiently complete and applicable to moral problems; instead, theory itself must be tested for adequacy by its practical implications. But the goal of reflective equilibrium protects against the the risk of prejudice and merely intuitive moral judgments, because theory and practice have a relation of reciprocity that is mutually constraining.

To conclude, we have agreed with Rawls that justification is "a matter of the mutual support of many considerations, of everything fitting together into one coherent whole."[25] We will now pursue this thesis by developing our own coherence-based account of justification.

A Coherence Theory of Justification

In the late 1970s, when the first edition of this book was published, we sketched an account of justification for practical ethics in response to a crisis then bedeviling the underdeveloped field of biomedical ethics. At that time there was no sustained theory of biomedical ethics and no systematic account of its principles and normative rules. As our editions have evolved, many readers have requested that we state our views on method and justification in greater depth. We do so in this section.

Following Joel Feinberg and certain traditions in Greek philosophy, we initially described the relation between moral experience and moral theories as *dialectical*. We develop theories to illuminate experience and to determine what we ought to do, but we also use experience to test, corroborate, and revise theories.[26] If a theory yields conclusions at odds with our considered judgments—for example, if it allows children but not adults to be used without consent as subjects of biomedical research—we have reason to be suspicious of the theory and to modify it or seek an alternative theory. We regard this dialectical strategy as a way to work toward coherence between particular and general judgments.[27] As Feinberg notes, this procedure is similar to the reasoning that occurs in courts of law. On the one hand, if a principle commits one to an antecedently unacceptable judgment in a particular case, then one should modify or supplement the principle so as to render it coherent with one's particular and general beliefs taken as a whole. On the other hand, when a well-founded principle indicates the need to change a particular judgment, the overriding claims of coherence require that the judgment be adjusted.[28] It seems mistaken, then, to say that principles are not *drawn from* cases but only *applied to* cases. Furthermore, both general and particular considered judgments provide data for theory and are theory's testing ground. They lead us to modify and refine embryonic theoretical claims, especially by pointing to inadequacies in or limitations of theories.[29]

Although justification is a matter of coherence, bare coherence has seemed to many philosophers an insufficient basis for justification, because the substantive body of judgments and principles that cohere could themselves be morally unsatisfactory. There also could be a series of alternative coherent systems, each with a claim as valid as the next if coherence alone is the judge of theories. Moral justification and knowledge could never be achieved unless some criterion independent of coherence were added to the account. An example of this problem appears in the so-called Pirates' Creed of Ethics or Custom of the Brothers of the Coast.[30] Formed under a democratic confraternity of marauders circa 1640, this creed for pirates is a coherent, carefully delineated set of rules governing mutual assistance in emergencies, penalties for prohibited acts, the distribution of spoils, modes of communication, compensation for injury, and "courts of honour" that resolve disputes. This substantive body of rules and principles, although coherent, is morally unsatisfactory. Its appeal to "spoils," its awarding of slaves as compensation for injury, and the like involve immoral activities. But what justifies us in saying this coherent code is not an acceptable code of ethics?

This question points to the importance of starting with considered judgments that are settled moral convictions in a broad expanse of ethics, and then casting the net more broadly in specifying, testing, and revising those convictions. Coherentism is not exhausted by a relentless reduction of any set of beliefs to coherence. We start in ethics, as elsewhere, with a particular set of beliefs—the set of considered judgments (also called self-evident norms and plausible intuitions) that are acceptable initially without argumentative support. We cannot justify every moral judgment in terms of another moral judgment without generating an infinite regress or vicious circle of justification in which no judgment is justified. The only avenue of escape is to accept some judgments as justified without recourse to other judgments; and these judgments form our starting point.

These considered judgments typically have a history rich in moral experience that underlies our sense that they are credible and trustworthy; considered judgments are therefore not merely a matter of individual intuition. Any moral certitude associated with these norms is likely to derive from beliefs that are acquired, tested, and modified over time in light of the purposes served by the norms. Coherence among these initial norms is essential to their acceptability, and incoherence is a sound reason for rejecting one or more such "foundational" but fallible propositions. The Pirates' Creed, while coherent and acceptable among pirates, fails the test of initial moral acceptability.

Although we start with initially credible premises, the persons, codes, institutions, or cultures from which the premises descend need not themselves be in every case highly reliable or comprehensive in their reports and documents.

For example, the Hippocratic tradition—the starting point in medical ethics for centuries—has turned out to be a limited and generally unreliable basis for medical ethics. This problem can be overcome in a coherence theory by calling on a wider body of experience to collect points of convergence. To use an analogy to eye-witnesses in a courtroom, if a sufficient number of entirely independent witnesses converge to agreement in recounting the facts of a story, the story gains credibility beyond the credibility of the individuals who tell it. At the same time, we can eliminate the witnesses' stories that do not converge and cannot be made consistent with the main lines of convergence. The greater the coherence in a broadly based account that descends from initially credible premises, the more likely we are to believe it. The same point holds in moral theory: As we increase the number of accounts, establish convergence, and increase coherence, the best explanation is that the beliefs are justified and should be accepted. When we find wider and wider confirmation of hypotheses, the best explanation is that these hypotheses are the right ones. Such confirmation is the proper goal of moral theory, however difficult it is to accomplish.

Does it follow that there are *degrees* of justification and knowledge? Often we can achieve only a weak coherence, using more or less reliable reports. The extent to which we can speak of beliefs as justified is comparative, contingent upon evidence and degree of coherence. Justification in areas such as abortion and animal rights is notoriously difficult and resistant to solution by the ideal of reflective equilibrium. We will hereafter assume, without further argument, that this degrees-of-justification thesis is correct. We will also assume that coherence is the central condition in moral justification, but not that it is the sole condition.

In addition, several safeguards should be recognized in attempts to reconstruct moral concepts and norms on the model of coherence. These safeguards function to protect against faulty coherence construction. One safeguard can be called *the resemblance condition*. It requires that a moral account remain faithful to (resemble) the principles and concepts that provided the starting point for that account. In selecting data and constructing a theory, the final product should resemble the principles and concepts that it explicates. For example, suppose we are attempting to develop a coherent account of medical confidentiality that will remove several serious problems in our current system of confidentiality in health care; and suppose we rely so heavily in our account on rights of privacy that we completely lose sight of medical confidentiality and supply only a list of rights of privacy that hospitals should acknowledge. If confidentiality has been lost in the process, we have failed to do what we set out to do, even if the product exhibits a high degree of coherence. Although room exists for disagreement over the proper departure point for such an analysis (and therefore over what the final product must resemble), resemblance is

an important constraint on theory construction. At the same time, resemblance must not be construed to preclude the possibility of radical shifts in normative perspective. The resemblance condition does not allow one principle or concept to be turned into another, but it also does not preclude our coming to the view that our initial judgment or position was flatly wrong.

Second, *universalizability* is a widely accepted condition that serves a safeguarding function. This condition does not imply that a society's distinct norms (those that differ from another society's) logically could not constitute a moral code or that all moral judgments and standards are identical for all persons—leaving no room for individual or group differences, for diverse moral traditions, for special relationships, and for autonomous judgments and moral disagreements. Rather, the condition of universalizability requires that any person who judges that action x is morally required (or morally worthy, morally virtuous, etc.) in circumstance C_1 is thereby committed to the premise that x is morally required (or morally worthy, morally virtuous, etc.) in circumstance C_2 if C_1 and C_2 are not different in any morally relevant respect. One therefore can universalize by advocating that all persons act in a certain way in a type of circumstance; but equally informed, rational, and impartial persons can advocate different actions in that type of circumstance. Universalizability is not a moral norm parallel to a material principle of justice or a demand for equal treatment; it is a formal condition rather than a substantive principle.

This condition implies that basic moral principles must be formulated in terms of universal rather than particular properties. Morality does not, for example, recognize a relevant difference between *I* and *he* or *she* in formulating norms of right or wrong.[31] This is one way in which morality protects against bias, prejudice, and idiosyncratic preference—at the same time recognizing that relevant differences exist between persons such as being a parent, being a supervisor, being experienced in a job, and the like. Universalizability, then, demands consistency of commitment within a moral system of judgments, rules, and principles, but it leaves open for discussion exactly what will count as morally relevant similarities and differences and whether unanimity is reachable over principles and rules.

Other safeguards also deserve mention. The comparative endurance, resilience, and output power of a principle or theory are clearly points in its favor. That a principle or theory *endures* through competitive encounters, is *adaptive to novelty*, and meets new problems with *creative and practical solutions* are all criteria of acceptability that promote reflective equilibrium. In Chapter 2 we develop a framework for theory construction that further supports the thesis that degrees of adequacy in theory are to be expected. In defending a coherence theory of justification, then, we do not mean to suggest that the criterion of coherence alone determines the merit of a theory.

Coherence and Noncoherence Accounts of Truth

It is customary in moral theory, epistemology, and philosophy of science to distinguish truth and justification. Problems such as the one encountered in our discussion of the Pirates' Creed suggest to some people that a moral theory must be *true*, not merely coherent. Others urge that a coherence theory of justification can double as a theory of truth. The idea is that convergence toward truth, not simply justification, is achieved by coherence (using an ideal such as reflective equilibrium). Still others deny this claim, saying that coherence is not constitutive of truth. They ask whether if several inconsistent theories generated from plausible initial premises can each claim a strong measure of internal coherence and consistency, we must then say that several true theories exist—for example, several true theories of justice.

This question leads us to ask whether a better criterion exists of the truth of a moral system than coherence among its norms. The problem with this proposal is that if a coherence theory is the right account of justification and no alternative route exists to the justification of a moral claim, there is no alternative way to check up on the results of a coherence account except coherence together with careful inspection of initial premises and of the safeguarding conditions mentioned above. To demonstrate the falsity of a belief, one would have to present a reasoned challenge to the results in the coherence account. But how would this challenge be mounted, except by casting more broadly the net of beliefs brought into coherence, which is part of the method of coherence?

The best explanation in the face of unshakable coherence is that the system has captured what is right, virtuous, and the like. If this result is what moral truth consists in, then the net of coherence has captured the truth. We would likely make such a claim in parallel cases of scientific knowledge: If a stable coherence is achieved after repeated testing, the best explanation is that the system of scientific beliefs so achieved either expresses or approximates the truth. If this is a sound account of scientific truth, is it not likewise a sound account of moral truth?

Yet it is far from settled that this treatment of moral truth is adequate. It is doubtful that a successful body of coherent beliefs, no matter how stable, yields truth. For one reason, it is doubtful that moral statements have truth values and that truth is a category that should appear in moral theory. For another reason, we would need a theory of *truth*, itself a complicated and controversial subject. We are content to conclude here that justification successfully occurs in ethics and that the right approach to justification is the coherence account we have outlined above and will augment in Chapter 2. If, in addition, some want to say that *justified* beliefs are *true* beliefs, we have no objection to this language,

but we make no such extended claim for the conclusions expressed in this volume, and we believe that such a claim is likely to produce more misunderstanding than illumination.

Specifying and Balancing Principles

We are now in a position to develop the methods of specification and balancing mentioned previously. These methods fill out the coherence model and provide strategies for the resolution of moral problems and the avoidance of intractable conflict.

Specification

Philosopher G. W. F. Hegel fittingly criticized Immanuel Kant for developing an "empty formalism" that preached obligation for obligation's sake, without any power to develop an "immanent doctrine of duties." Hegel thought all "content and specification" in a living code of ethics had been replaced by abstractness in Kant's account.[32] Ethical theory that features *principles* has been similarly accused.[33] These criticisms point to an important problem. Every general norm, indeed morality itself, contains regions of indeterminacy that need reduction through further development and enrichment. If the principles discussed in this book are to have sufficient content, we must be able to specify the content in a way that surpasses ethereal abstractness, while also indicating the cases that properly fall under the principles.[34] If a principle lacks adequate specificity, it is empty and ineffectual.

Consider a simple example of this problem. Nonmaleficence is the principle that we ought not to inflict evil or harm on others. This principle provides only a rough starting point for guidance about the conditions under which harmful actions are prohibited. Normally we regard causing someone's death as a harm, but are assisted suicide and voluntary active euthanasia *harmful* actions that are absolutely proscribed by the principle of nonmaleficence? Are acts of mercy killing themselves sometimes acts of nonmaleficence, or even beneficence? If we question whether a physician who helps a patient commit suicide thereby harms or benefits the patient, no guidance is forthcoming from an unspecified principle of nonmaleficence. Without further specification, nonmaleficence is a bare starting point for resolving such problems as assisted suicide and euthanasia.

Abstract principles, then, often must be developed conceptually and shaped normatively to connect with concrete action-guides and practical judgments. In tightening our principles, we must take into account various factors such as efficiency, institutional rules, law, and clientele acceptance. Eventually we need to provide a practical strategy for real-world problems involving the demands of political procedures, legal constraints, uncertainty about risk, and the like. In light of the indeterminacy inherent in general norms, we accept Henry Richardson's argument that the specification of our principles is essential to determining what counts as an instance of that principle and to overcome some moral conflicts. Richardson notes that we sometimes *apply* norms directly to cases and that we often try to *balance* conflicting norms. Both techniques work on some occasions. But in managing new, complex, or problematic cases, the first line of attack should be to specify our norms and thereby to specify unclarities and problems away. In difficult cases, direct application rarely works, whereas balancing often appears to be too subjective, and fails to reduce conflict or the potential for further conflict. Specification, then, is an attractive strategy for the hard cases as long as the specification can be justified.[35] Of course, many already specified rules will need further specification to handle new circumstances of conflict. Progressive specification often must take place to handle the variety of problems that arise, gradually reducing the dilemmas and circumstances of conflict that the abstract principle has insufficient content to resolve.

As a simple example of specification, consider again the rule "Doctors should put their patients' interests first." A fact of life in modern medicine in the United States is that patients sometimes can afford the best treatment strategy only if their physicians falsify information on insurance forms, or at least only thinly spread the truth. It does not follow from a proper understanding of the rule of patient-priority that a physician should act illegally by lying or distorting the description of a patient's problem on an insurance form. Our rules against deception and for patient-priority are not categorical demands, and they stand in need of specification to give fuller, more concrete moral advice to physicians who wonder whether they should deceive payers, and, if so, under which conditions.

Nonetheless, the specification of a physicians' commitment to patients and to nondeception faces many problems, as a recent survey of practicing physicians' attitudes toward deception illustrates. Dennis H. Novack and several colleagues used a questionnaire to obtain physicians' responses to four difficult ethical problems that potentially could be resolved by deception. In one scenario, a physician recommends an annual screening mammography for a 52-year-old woman who protests that last year her insurance company would not cover the test and she had to pay herself, although she could not afford it. A secretary suggests that the patient's insurance company would cover the costs of the mammography if the physician stated the reason as "rule out cancer" rather than "screening mammography," although the latter alone was the reason. Almost seventy percent of the physicians responding to this survey indicated that they would put "rule out cancer," and eighty-five percent of this group insisted that their act would not involve "deception."[36]

These physicians' decisions can be interpreted as crude attempts to specify rules against deception. Most physicians in the study apparently did not operate with the definition of deception favored by the researchers ("to deceive is to make another believe what is not true, to mislead"). Perhaps the physicians believed that deception involves withholding information from or misleading someone who has a *right* to that information, and also believed that an insurance company with unjust policies of coverage has no right to accurate information. Or perhaps they believed that "deception" occurs when one unjustifiably misleads another, and that it was justifiable to mislead the insurance company in these circumstances. Yet another possibility is that these physicians understood the rule against deception to prohibit only self-serving actions or actions that harm other individuals, or to permit deceptive actions on behalf of one's patients whose interests should come first.

These physicians would not agree on how to specify rules against deception as well as rules requiring that patients' interests be first. Each of the proposed specifications would resolve the conflict (or, perhaps better, would *dissolve* it), but there would be debate about the justifiability of each specification. This survey provides an example of eliminating an apparent dilemma without either "applying" or "balancing" norms, yet any *proposed* specification may fail to provide the most adequate or justified resolution. To say that a problem or conflict is "resolved" or "dissolved" is here only to say that norms have been made determinate in content so that, when cases fall under them, we can decide what ought to be done.

An adequate specification requires that one justify the claim that the proposed specification is coherent with other relevant moral norms. Specification is a way of resolving problems through deliberation, but no proposed specification is justified without a showing of coherence. All moral norms are, in principle, subject to such revision, specification, and justification. The reason for this constant need for further content, as Richardson puts it, is that "the complexity of the moral phenomena always outruns our ability to capture them in general norms."[37] These problems about specification do not undercut the practical point that we can sometimes satisfactorily resolve conflicts and dilemmas by adequate specification.

Whether our framework is a workable account for biomedical ethics depends in part on whether its principles and related rules can be specified and the specifications justified. The substantive rules, authority rules, and procedural rules discussed below (pp. 38–39) involve specifications of our framework principles, and much of the argument of Chapters 3–7 is intended to show this process in action.

However, we still need to note some limitations and weaknesses in the method of specification, to make clear that we do not view the method as a cure-all for our deepest dilemmas. First, opposition between the model of

specification and the models of balancing and applying norms should not be overstated. Nothing in the model of specification indicates that there is a way around balancing in the very act of specifying principles and rules; and nothing in the model shows that straightforward application never works. In any given problematic or dilemmatic case, several competing specifications will typically constitute possible resolutions, thereby returning us to conflicts of the sort that drove us to specification in the first place. (We explicate the model of balancing below.) Second, if one believes, as we do, that some moral conflict is inevitable and cannot always be avoided or eliminated by even tightly knit specifications, then the method is suited only for contexts in which specification has a reasonable hope of acceptance. Third, making norms more specific does not itself preclude the use of dogmatic, biased, arbitrary, or irrational views to make one's favored conclusion correct by fiat. Even if a specification eliminates contingent conflict, the specification may be arbitrary, lack impartiality, or fail for other reasons. As Richardson forthrightly acknowledges, "once the operation of specification has been adequately understood, it may then be admitted that it should be supplemented by application and balancing in a more complex hybrid model."[38]

Specification as a method must be indissolubly connected with a larger model of coherence that appeals to considered judgments and to the overall coherence introduced by a proposed specification. This is a general model that we accept, as does Richardson (who holds that specification and the coherence ideal dovetail, but have distinctive roles). So understood, specification holds out the possibility of a continually expanding normative viewpoint that is faithful to initial beliefs (which are not renounced) and that tightens rather than weakens coherence among the full range of accepted norms.

What advantage, then, does specification have over other attempts to resolve problems? Richardson's response, like ours, is that this question should be "answered in terms of the overall coherence and mutual support of the whole set of moral norms. . . . A coherence standard for the rationality of specification . . . in effect carries the Rawlsian idea of 'wide reflective equilibrium' down to the level of concrete cases."[39] From this perspective, specification is one arm of a larger method of coherence—a view that reinforces our earlier arguments that the central condition of justification is coherence and that interpretation, construction, and reconstruction are essential in both ethical theory and practical ethics. A particular specification is justified only if it is more coherent with the whole set of relevant norms than any other available specification.

The upshot of our analysis of coherence and specification is the following: One goal of a moral theory, and central to its account of justification, is to move from general levels of theory to particular rules, judgments, and policies that are in close proximity to everyday decisions in the moral life. Like a

tributary with many forks into different territories, the principles and rules in a theory can be made to fork outward through specification and feed different parts of the moral life. Appropriate specification conserves or elevates the coherence already present in the theory. When moral conflicts occur, specification supplies an *ideal* of repeated coherence testing and modification of a principle or rule until the conflict is specified away, but specification is also a useful tool for the development of *policies* in biomedical ethics.

To accept this ideal is not to assume that conflicts can always be specified away by developing rules or policies. The moral life will be plagued by contingent conflicts that cannot be eliminated. Our pragmatic goal should be a method of resolution that often helps, not a method that will invariably resolve our problems.

Balancing and Overriding

Principles, rules, and rights require *balancing* no less than *specification*. Principles (and the like) direct us to certain forms of conduct, but principles by themselves do not settle conflicts of principle. Whereas specification entails a substantive development of the meaning and scope of norms, balancing consists of deliberation and judgment about the relative weights of norms. Balancing sometimes occurs in specification, and specification also sometimes occurs in balancing. Specification and balancing can best be conceived as mutually facilitative approaches, methods, or strategies that fit coherently within the larger method of coherence outlined above. Balancing is especially useful for individual cases, whereas specification is especially useful for policy development.

Avoiding balancing through "absolute" norms. Throughout this book we view the norms to be balanced—principles, rules, rights, and the like—as *prima facie* (see below, pp. 33–37), and not as absolute, as rules of thumb, or as hierarchically (lexically or serially) ordered. However, some specified norms are virtually absolute, and therefore usually escape the need to balance. Examples include prohibitions of cruelty and torture, where these actions are defined as gratuitous infliction of pain and suffering. (Other prohibitions, such as rules against murder, are absolute only because of the meaning of their terms. For example, to say "murder is categorically wrong" is to say "unjustified killing is unjustified.")

Defensible substantive absolutes are thorough and decisive specifications of principles. They are rare and rarely play a role in moral controversy. More interesting are norms that are formulated with the goal of including all legitimate exceptions, but whose formulation remains controversial. An example is "Always obtain oral or written informed consent for medical interventions with competent patients, *except* in emergencies, in low-risk situations, or when patients have waived their right to adequate information." This norm clearly needs interpretation and a specification of what constitutes an informed consent, an emergency, a waiver, and a low risk, but this rule would be absolute if all legitimate exceptions were included in its formulation and specification. Depending on how far specification goes and whether all exceptions are built in and defended, a rule could turn out to be legitimately absolute and thereby escape balancing because its potential for conflict with other principles and rules would have been eliminated. However, if such rules exist, they are rare. Moreover, in light of the enormous range of possibilities for contingent conflicts among rules, absolute rules are best construed as ideals rather than finished products.

Balancing prima facie norms. Drawing on W. D. Ross, we distinguish *prima facie* obligations from *actual* obligations. *Prima facie obligation* indicates an obligation that must be fulfilled unless it conflicts on a particular occasion with an equal or stronger obligation. A prima facie obligation is binding unless overridden or outweighed by competing moral obligations. Acts often have several morally relevant properties or consequences. For example, an act of lying may also promote someone's welfare, and an act of killing may involve the relief of pain and suffering as well as respect for a patient's autonomy through causing the patient's death at the patient's request. Such acts are at once prima facie wrong and prima facie right, because two or more norms conflict in the circumstances. The agent must then determine what he or she ought to do by finding an actual or overriding (in contrast to prima facie) obligation; that is, the agent must locate what Ross called "the greatest balance" of right over wrong. An agent's actual obligation in the situation is determined by the balance of the respective weights of the competing prima facie obligations (the relative weights of all competing prima facie norms such as beneficence, fidelity, and justice).[40] This metaphor of larger and smaller weights moving a scale up and down graphically depicts the balancing process, but it may also obscure what happens in the process of balancing by misleadingly suggesting an intuitive or subjective assessment. Justified acts of balancing entail that good reasons be provided for one's judgment.

For example, suppose a physician encounters an emergency case that would require her to extend an already long day so that she would be unable to keep a promise to take her son to the library. She will then engage in a process of deliberation that leads her to consider how urgently her son needs the visit to the library, whether they could go very late to the library, whether another physician could handle the case, etc. If she determines to stay deep into the night with the patient, this obligation will have become overriding because she has a good and sufficient reason. A life hangs in the balance, and she alone has the knowledge to deal adequately with the full array of the circumstances.

Her action of canceling her evening with her son, painful and distressing as it is, can be justified by this good and sufficient reason for doing what she does. Balancing, then, is a process of justification only if adequate reasons are presented.

One way of viewing this process brings it close to and perhaps merges it with specification. As David DeGrazia has pointed out to us, the good and sufficient reasons that one offers in *an act of balancing* can be viewed as a *specification of norms that incorporates one's reasons*. These reasons can be generalized for similar cases: "If a patient's life hangs in the balance and the attending physician alone has the knowledge to deal adequately with the full array of the circumstances, then the physician's conflicting domestic obligations must yield." This merging of specification and balancing has merits, but it may be too neat and too sweeping to handle all situations of balancing and specification. Balancing often eventuates in specification, but it need not; and specification often involves balancing, but it also might only add details. Accordingly, we do not propose to merge the two methods. The point is that balancing does not compete with specification, and they both coherently augment the model of coherence. We therefore propose that balancing and specification be seamlessly united with a general model of coherence that requires us to defend the reasons we give for actions and norms. As noted previously, balancing is particularly useful for case analysis, and specification for policy development.

Conditions that restrict balancing. As a response to criticisms that the model of balancing is too intuitive and open-ended, we can list a few minimal conditions that reduce the amount of intuition involved. These conditions add content to the requirement of giving good reasons for actions and norms. The following conditions must be met to justify infringing one prima facie norm in order to adhere to another (however, these conditions, being norms themselves, are also prima facie, not absolute):

1. Better reasons can be offered to act on the overriding norm than on the infringed norm (for example, typically if persons have a *right*, their interests deserve a special place when balancing those interests against the interests of persons who have no comparable right).
2. The moral objective justifying the infringement has a realistic prospect of achievement.
3. No morally preferable alternative actions can be substituted.
4. The form of infringement selected is the least possible, commensurate with achieving the primary goal of the action.
5. The agent seeks to minimize the negative effects of the infringement.

Although some of these conditions appear to be tautological, or at least entirely noncontroversial, in our experience they often are not observed in moral deliberation and would lead to different actions were they observed. For example, many proposals in biomedical ethics about the use of life-extending technologies seem to violate (2) by endorsing certain actions in which no realistic prospect exists of achieving the goals of a proposed medical intervention. Typically this occurs when the intervention is regarded by the health professionals as legally required, but in some cases the intervention occurs merely as a matter of routine practice. Even more commonly violated is condition (3). Actions are frequently performed without serious consideration being given to the range of preferable alternative actions that might be taken when one obligation is in conflict with another obligation. For example, in animal care and use committees, a common conflict involves the obligation to approve a good scientific protocol and the obligation to protect animals against unnecessary suffering. The protocol is typically approved if a standard form of anaesthesia is proposed in the protocol. But standard forms of anaesthesia are often not the best way to protect the animal, and further inquiry is needed to determine the best anaesthetic for the interventions proposed. In our schema of conditions, it is *unjustifiable* to approve the protocol or to conduct the experiment without this additional inquiry. Accordingly, we think the above conditions are morally demanding, not simply obvious or tautological. When conjoined with our requirements of coherence, these minimal conditions should help us achieve some measure of protection against arbitrary, purely intuitive judgments.

But even with these safeguards, controversy will arise regarding which norm should triumph in a particular conflict with another norm. We could try to introduce further criteria or safeguards, such as, "rights override non-rights" and "liberty principles override nonliberty principles," but these meta-rules are certain to fail in many circumstances in which rights claims and liberty interests are relatively minor. Honesty about the process of balancing and overriding compels us to return to our earlier discussion of dilemmas and to acknowledge that in some circumstances we will not be able to determine which moral norm is overriding. Balancing is further complicated by the wide range of relevant considerations. Sometimes we must consider matters such as whether a personal relationship with a long history gives special weight to one party's interests or claims, whether one party likely to suffer a loss can be compensated for the loss whereas another party cannot be compensated, and whether being responsible for causing a harm gives the party harmed a special claim that another party does not have. To illustrate the last circumstance, suppose X causes a harmful outcome to Y. We typically think Y has a greater claim for assistance or compensation from X than does Z who, like Y, has suffered a harmful outcome, but not one caused by X; Y's interests usually deserve more weight in X's balancing even if Z has suffered the greater harm.

In all of these cases some intuitive judgments and subjective weightings are unavoidable, just as they are everywhere in life when we must balance competing goods (for example, in the foods we eat, in the strategies we try in tennis, and in the way we allocate time in our daily schedules). But this fact does not reduce the process of balancing and overriding to arbitrary or merely subjective preferences. Consider a typical example. The principle of respect for autonomy and the principle of beneficence (which requires acts of preventing harm to others) sometimes conflict in the AIDS epidemic. Respect for autonomy sets a prima facie barrier to the mandatory testing of people who are at risk of HIV infection and whose actions may put others at risk, and yet society has a prima facie obligation to act to prevent harm to those at risk. The two prima facie principles conflict, but to justify overriding respect for autonomy, one must show that the mandatory testing of certain individuals is necessary to prevent harm and has a reasonable prospect of preventing harm. If it meets these conditions, mandatory testing will still need to pass the least-infringement test, and agents must seek to reduce the harmful effects (such as the negative consequences that individuals fear from testing). As we will see in Chapter 7, many (but not all) proposed forms of mandatory testing cannot be justified, because other available alternatives would have a higher probability of success without infringing personal autonomy.[41]

As with specification, the process of balancing cannot be rigidly dictated by some formulaic "method" in ethical theory. The model of balancing will satisfy neither those who seek clear-cut, specific guidance about what one ought to do in particular cases nor those who believe in a lexical or serial ranking of principles, with automatic overriding conditions. Some have therefore attempted to circumvent balancing judgments by delineating automatic-overriding features in their ethical theories. For example, some have spoken of rights as *trumps* (Ronald Dworkin) and as *side constraints* on permissible judgments (Robert Nozick). But these attempts have failed because the proposed trumps and side constraints themselves must be balanced in many circumstances (see Chapter 2, pp. 71–72).

Ross's distinction between prima facie and actual obligations, as well as his model of balancing, are also attractive in that they conform closely with our experience as moral agents. The lingering concern about the role of intuition and subjectivity, even in the context of giving good reasons, does not disqualify the model. We can reflect on troublesome moral problems even if plural and conflicting values make comparisons difficult. A plurality of values and judgments does not by itself stifle sound deliberation, balancing, justification, and decisionmaking. Almost daily we are confronted with situations in which we must make choices between plural and conflicting values in our personal lives, and we must balance several considerations. Some of these choices are moral, but many are nonmoral. For example, our budget may require that we make a choice between buying books or buying a train ticket to see our parents. Not having the books will be an inconvenience and a loss, and not visiting home will make our parents unhappy. The choice is not easy, perhaps, but usually we think through the alternatives, deliberate, balance, and reach a conclusion.

Our concern in this section has not been to produce a mechanistic method for balancing, any more than we produced a definitive method of specification in the preceding section. We have proposed that a model of balancing deserves the same serious consideration as the model of specification and that both are needed for an account of moral judgment. In reaching our conclusions about method and justification, we have yet to defend any normative premises or principles. In the remainder of this chapter, we will outline (but not yet defend) the structural features of the normative framework adopted in this book.

The Place of Principles

We defend what has sometimes been called the *four-principles approach* to biomedical ethics,[42] and also called, somewhat disparagingly, *principlism*.[43] These principles initially derive from considered judgments in the common morality and medical tradition that form our starting point in this volume. For example, the principle of beneficence derives, in part, from long-standing, professional role obligations in medicine to provide medical benefits to patients. Our goal is to specify and balance these principles by the methods of ethical theory previously discussed. Both the set of principles and the content ascribed to the principles are based on our attempts to put the common morality as a whole into a coherent package.

In this section we sketch our ethical framework by providing an analytical breakdown of its elements. We distinguish several types of normative action-guides as components in our framework, including principles, rules, rights, and virtues. Although rules, rights, and virtues are of the highest importance for health care ethics, principles provide the most abstract and comprehensive norms in the framework. These principles will be individually analyzed in later chapters. This chapter presents only the structural shell.

Four Clusters of Basic Principles

We begin with our assumptions. That four clusters of moral "principles" (in another framework they might be developed as "rights," "virtues," or "values") are central to biomedical ethics is a conclusion we have reached by the search for considered judgments and coherence, not a position that receives an argued defense. However, we will in later chapters defend the choice of each principle as well as the independent significance of each. We also operate with

only a loose distinction between rules and principles. Both are normative generalizations that guide actions, but, as we analyze them, *rules* are more specific in content and more restricted in scope than principles. Principles do not function as precise action guides that inform us in each circumstance how to act in the way more detailed rules do. *Principles* are general guides that leave considerable room for judgment in specific cases and that provide substantive guidance for the development of more detailed rules and policies. This limitation is no defect in principles; rather, it is a part of the moral life in which we are expected to take responsibility for the way we bring principles to bear in our judgments about particular cases. We also need to distinguish both rules and principles from the coherent, systematic body of norms that comprise *theories*. Our discussion of the coherence theory has already given some insight into our views about the nature of theory and its construction, but we defer further discussion until Chapter 2.

The four clusters of principles are (1) respect for autonomy (a norm of respecting the decisionmaking capacities of autonomous persons), (2) nonmaleficence (a norm of avoiding the causation of harm), (3) beneficence (a group of norms for providing benefits and balancing benefits against risks and costs), and (4) justice (a group of norms for distributing benefits, risks, and costs fairly). Nonmaleficence and beneficence have played a central historical role in medical ethics, whereas respect for autonomy and justice were neglected in traditional medical ethics but have came into prominence because of recent developments.

To illustrate this point about historical significance and neglect, British physician Thomas Percival furnished our first well-formed doctrine of medical ethics in 1803. His work served as the prototype for the American Medical Association's (AMA) first code of ethics in 1847. Easily the dominant influence in both British and American medical ethics of the period, Percival argued (using somewhat different language) that nonmaleficence and beneficence fix the physician's primary obligations and triumph over the patient's preferences and rights in any circumstance of serious conflict.[44] Percival failed to foresee the power of principles of respect for autonomy and distributive justice, but in fairness to him, it should be acknowledged that considerations of respect for autonomy and distributive justice are now ubiquitous in discussions of biomedical ethics in a way they were not when he wrote in late eighteenth century Britain.

Types of Rules

In addition to the four clusters of principles, we defend several types of rules that specify principles and guide actions.

Substantive rules. Rules of truthtelling, confidentiality, privacy, fidelity, and various rules pertaining to the allocation and rationing of health care, omitting treatment, physician-assisted suicide, and informed consent need to be formulated as guides to action that are more specific than abstract principles. A typical example of a rule that specifies the principle of respect for autonomy by giving it more content is, "Follow a patient's advance directive whenever it is clear and relevant."

Authority rules. We also defend rules about decisional authority—that is, rules regarding *who* may and should perform actions. For example, *rules of surrogate authority* determine who should serve as surrogate agents in making decisions for incompetent persons, and *rules of professional authority* determine who, if anyone, should make decisions to override or to accept a patient's decisions if they are medically damaging and poorly considered. Another example is found in *rules of distributional authority* that determine who should make decisions about the allocation of scarce medical resources. These authority rules do not delineate substantive standards or criteria for making decisions. Substantive standards—such as *guidance rules* for surrogate decisionmaking (advance directives, substituted judgment, and best interests) and *rationing rules* for the allocation of scarce resources (such as constituency priority and medical utility)—are moral directives that belong in the first category of substantive rules. Although authority rules are distinct in type from substantive rules, they interact in both theory and practice. For instance, authority rules are justified in part by how well they express substantive rules and principles.

Procedural rules. We also defend rules that establish procedures to be followed. Procedures for determining eligibility for scarce medical resources and procedures for reporting grievances to higher authorities are typical examples. We often resort to procedural rules when we run out of substantive rules and when authority rules are incomplete or inconclusive. For example, if substantive or authority rules are inadequate to determine which patients should be given scarce medical resources, we resort to procedural rules such as first-come-first-served, queuing, and lottery. (See pp. 382–384 in Chapter 6.)

Rights, Virtues, Emotions, and Assorted Moral Considerations

Our framework of principles and rules, as expressed above, does not specifically incorporate the rights of persons, the character and virtues of the agents who perform actions, or the moral emotions. These moral considerations all merit attention in a comprehensive theory. Rights, virtues, and emotional responses are in some contexts of greater moral importance than principles and rules. For example, an ethics of virtue helps us see why good moral choices

depend on more than principles, and it also allows us to assess a person's moral character in a richer way than an ethics of principles and rules does. At this point we note only two points about our basic framework. First, it does not exclude categories such as rights, emotions, and virtues; we will in due course incorporate these categories. Second, we believe that for biomedical ethics, which has concentrated on guidelines for action, principles and rules are both indispensable and central to the enterprise. However, we will qualify this conclusion in Chapters 2 and 8, in which we clarify the role of these various categories.

Conclusion

In this chapter we have explained why moral reasoning is more complicated than the outmoded label "applied ethics" suggests. We have also hinted at an interdisciplinary account of biomedical ethics. Any solidly grounded discipline of ethics involves obtaining relevant factual information, assessing its reliability, and mapping out alternative solutions to problems that have been identified. This mapping sometimes entails presenting and defending reasons in support of factual, conceptual, and moral claims, while at the same time analyzing and assessing basic assumptions and commitments. Ethical theory, then, is but one vital contributor among other disciplines, including medicine, nursing, public health, law, and the social sciences.

We limited our discussion of theory in this chapter primarily to questions of method, deliberation, justification, and truth. These problems belong to metaethics. Except for a brief sketch of our normative framework, we have avoided discussion of types of normative theory that stand to make a substantive contribution to biomedical ethics. That topic is the subject of our next chapter.

Notes

1. For discussion of whether a sharp distinction can be drawn between metaethics and normative ethics, see David O. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989).
2. Talcott Parsons, *Essays in Sociological Theory*, Rev. Ed. (Glencoe, IL: The Free Press, 1954), p. 372.
3. The American Medical Association Code of Ethics of 1847, largely adapted from Thomas Percival's *Medical Ethics; or a Code of Institutes and Precepts, Adapted to the Professional Conduct of Physicians and Surgeons* (Manchester: S. Russell, 1803), was a response to a crisis in public and professional confidence. See Donald E. Konold, *A History of American Medical Ethics 1847–1912* (Madison: State Historical Society of Wisconsin, 1962), ch. 1–3; and Chester Burns, "Reciprocity in the Development of Anglo-American Medical Ethics," in *Legacies in Medical Ethics*, ed. Chester Burns (New York: Science History Publications, 1977).

4. See Chapter 2 for a discussion of "The Patient's Bill of Rights." For its history, see Ruth R. Faden and Tom L. Beauchamp, *A History and Theory of Informed Consent* (New York: Oxford University Press, 1986), ch. 3.
5. Jay Katz, ed., *Experimentation with Human Beings* (New York: Russell Sage Foundation, 1972), pp. ix–x.
6. See the publications of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research and of the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, several of which are mentioned throughout this volume.
7. Omnibus Budget Reconciliation Act of 1990. Public Law 101–508 (Nov. 5, 1990), §§ 4206, 4751. See 42 USC, scattered sections.
8. See Dennis Thompson, "Philosophy and Policy," *Philosophy and Public Affairs* 14 (Spring 1985): 205–18; and a Symposium on "The Role of Philosophers in the Public Policy Process: A View from the President's Commission," with essays by Alan Weisbard and Dan Brock, in *Ethics* 97 (July 1987): 775–95.
9. See John Lemmon, "Moral Dilemmas," *Philosophical Review* 71 (1962): 139–58.
10. William R. Bascom, *African Dilemma Tales* (The Hague: Mouton, 1975), p. 145 (relying on anthropological research by Roland Fletcher).
11. See the essays in Christopher W. Gowans, ed., *Moral Dilemmas* (New York: Oxford University Press, 1987); see also Walter Sinnot-Armstrong, *Moral Dilemmas* (Oxford: Basil Blackwell, 1988) and Edmund N. Santurri, *Perplexity in the Moral Life: Philosophical and Theological Considerations* (Charlottesville: University Press of Virginia, 1987).
12. John Dewey, *Theory of the Moral Life* (New York: Holt, Rinehart and Winston, 1960), p. 135, and Dewey and James H. Tufts, *Ethics* (New York: Henry Holt and Co., 1908), p. 323.
13. *United Automobile Workers v. Johnson Controls, Inc.*, Slip opinion (Argued October 10, 1990—Decided March 20, 1991).
14. One who defends this proposition might wish to stipulate that it is possible for the doctor to perform the action, that the resources can be obtained, that no competing obligation is overriding in the circumstances, and so forth. However, the precise nature of a fully specified principle is independent of our present inquiry.
15. J. B. Schneewind, "Moral Knowledge and Moral Principles," in *Revisions: Changing Perspectives in Moral Philosophy*, ed. Stanley Hauerwas and Alasdair MacIntyre (Notre Dame, IN: University of Notre Dame Press, 1983), pp. 118–20.
16. Henry Sidgwick, *The Methods of Ethics* (Indianapolis, IN: Hackett Publishing Co., 1981), bk. I, ch. 8, § 2, p. 99.
17. K. Danner Clouser and Bernard Gert, "A Critique of Principlism," *The Journal of Medicine and Philosophy* 15 (April 1990): 232.
18. Albert R. Jonsen and Stephen Toulmin, *The Abuse of Casuistry* (Berkeley: University of California Press, 1988), p. 16.
19. Albert R. Jonsen, "Of Balloons and Bicycles or the Relationship between Ethical Theory and Practical Judgment," *Hastings Center Report* 21 (September–October 1991): 14–16.
20. Stephen Toulmin, "The Tyranny of Principles," *Hastings Center Report* 11 (December 1981): 31–39.
21. John Rawls, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 48 (1974–75): 8.
22. Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), pp. 20–21, 46–50, 579–80.

23. See Norman Daniels, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76 (May, 1979): 257ff. Henry Richardson has pointed out to us that Rawls does not make it a logically necessary condition of considered judgments that they be *shared* with others; nor does he make it a logically necessary condition of wide reflective equilibrium that it is shared. However, to make his enterprise of social justice work, the considered judgments he selects would have to be widely shared. The importance of shared agreement is brought out in Rawls's emphasis on his theory of justice as a liberal political theory. (Richardson also argues that Rawls's primary protection against local bias is in his account of the veil of ignorance rather than in wide reflective equilibrium. Of course one could argue that both serve this purpose.)
24. Compare Rawls, *A Theory of Justice*, pp. 195–201.
25. *A Theory of Justice*, pp. 21, 579.
26. *Principles of Biomedical Ethics*, 1st Ed. (1979), esp. pp. 13–14. Work in biomedical ethics in the 1970s consisted almost entirely of articles and essays. Various problem areas were treated, such as abortion, euthanasia, and the allocation of resources. The few systematic statements and their discussions were arranged in terms of these problem areas. Examples include Joseph Fletcher, *Morals and Medicine* (Princeton, NJ: Princeton University Press, 1954), Paul Ramsey, *The Patient as Person* (New Haven, CT: Yale University Press, 1970), and Howard Brody, *Ethical Decisions in Medicine* (Boston: Little, Brown and Company, 1976).
27. This perspective is suggested by Aristotle's discussion of dialectic as a means to first principles in the *Topics*; see *The Complete Works of Aristotle*, ed. Jonathan Barnes, vol. I (Princeton, NJ: Princeton University Press, 1984), 101a25–101b4.
28. Joel Feinberg, *Social Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1973), pp. 34–35. Chaim Perelman's account also influenced us in our second edition (1983): "In morals absolute preeminence cannot be given either to principles—which would make morals a deductive discipline—or to particular cases—which would make it an inductive discipline. Instead, judgments regarding particulars are compared with principles and preference is given to one or the other according to a decision that is reached by resorting to the techniques of justification and argumentation." *The New Rhetoric and Humanities: Essays on Rhetoric and Its Applications* (Boston: D. Reidel, 1979), p. 33.
29. See Judith Jarvis Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge, MA: Harvard University Press, 1986), pp. 251–60.
30. Circa 1640. Published 1974 by Historical Documents Co.
31. See R. M. Hare, *Moral Thinking: Its Levels, Method and Point* (Oxford: Clarendon Press, 1981), p. 223. But see also proposed qualifications on this view introduced in our Chapter 2, especially in the discussion of the ethics of care.
32. G. W. F. Hegel, *Philosophy of Right*, T. M. Knox, trans. (Oxford: Clarendon Press, 1942), pp. 89–90, 106–7.
33. Clouser and Gert, "A Critique of Principlism," Ronald M. Green "Method in Bioethics: A Troubled Assessment," *The Journal of Medicine and Philosophy* 15 (1990): 179–97, and Stephen Toulmin, "The Tyranny of Principles."
34. As R. M. Hare notes, "any attempt to give content to a principle involves specifying the cases that are to fall under it. . . . Any principle, then, which has content goes some way down the path of specificity." *Essays in Ethical Theory* (Oxford: Clarendon Press, 1989), p. 54.
35. Henry S. Richardson, "Specifying Norms as a Way to Resolve Concrete Ethical Problems," *Philosophy and Public Affairs* 19 (Fall 1990): 279–310. See also David DeGrazia's reflections on specification in "Moving Forward in Bioethical Theory: Theories, Cases, and Specified Principlism," *Journal of Medicine and Philosophy* 17 (1992): 511–39.
36. Dennis H. Novack, et al., "Physicians' Attitudes Toward Using Deception to Resolve Difficult Ethical Problems," *Journal of the American Medical Association* 261 (May 26, 1989): 2980–85.
37. Richardson, "Specifying Norms," p. 294. ("Always" in this formulation should perhaps be understood to mean "in principle always"; specification may, in some cases, reach a final form.) For an example of elementary specification (but not so called) using the four-principles approach, see Raanan Gillon, "Doctors and Patients," *British Medical Journal* 292 (1986): 466–69.
38. Richardson, "Specifying Norms," p. 280.
39. Ibid., pp. 299–300.
40. See W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), esp. pp. 19–36; and *The Foundations of Ethics* (Oxford: Clarendon Press, 1939).
41. See James F. Childress, "Mandatory Screening and Testing," in *AIDS and Ethics*, ed. Frederic G. Reamer (New York: Columbia University Press, 1991), pp. 50–76. For a sensitive attempt to balance the rights and interests of HIV infected surgeons and dentists against the rights and interests of their patients—an attempt that reaches conclusions similar to ours about balancing and overriding—see Norman Daniels, "HIV-infected Health Care Professionals: Public Threat or Public Sacrifice?," *The Milbank Quarterly* 70 (1992): 3–42, esp. 26–32.
42. See, for example, *Principles of Health Care Ethics*, ed. Raanan Gillon and Ann Lloyd (London: John Wiley & Sons, 1993).
43. Clouser and Gert, "A Critique of Principlism," pp. 219–36.
44. Thomas Percival, *Medical Ethics*. See note 4 above.

2

Types of Ethical Theory

A well-developed ethical theory provides a framework within which agents can reflect on the acceptability of actions and can evaluate moral judgments and moral character. This chapter concentrates on several types of ethical theory: utilitarianism, Kantianism, character ethics, liberal individualism, communitarianism, the ethics of care, casuistry, and common-morality accounts. Some knowledge of these theories is indispensable for reflective study in biomedical ethics because much of the field's literature draws on methods and conclusions found in these theories.

A conventional introduction to ethical theory explicates and then proceeds to criticize several leading ethical theories. Typically, the criticisms are so harsh that each theory seems demolished beyond repair. As a result, readers become skeptical about the value of ethical theory. This outcome is both unfortunate and unnecessary. Although defects and excesses appear in all theories, several contain insightful perspectives and compelling arguments. Only such theories are discussed in this chapter. Our goal is to eliminate what is unacceptable in these theories and to appropriate what is relevant and acceptable.

Occasionally we refer to our system and to arguments in this book as a *theory*. A word of caution is in order about this use of the word *theory*. This term is commonly used in ethics to refer to each of the following: (1) abstract reflection and argument, (2) systematic reflection and argument, and (3) an integrated body of principles that are coherent and well developed (in ways discussed in Chapter 1). We have attempted in this book to construct a coherent

account adequate for the particular subject of biomedical ethics, but we do not claim to have developed or to presuppose any particular comprehensive ethical theory in ways suggested by (3). We engage *in theory* (for example, in evaluating other ethical theories), and so in abstract reflection and argument (1). We also present an organized system of principles, and so engage in systematic reflection and argument (2). But, at best, we present only some elements of a comprehensive *general* theory (3). Our relevant presuppositions and theses are presented in the section near the end of this chapter titled "Principle-Based, Common-Morality Theories."

Each section of this chapter, except the first and the last, is divided into subsections, as follows: (1) an overview of the characteristic features of the theory (introduced by examining how its proponents would approach a case); (2) a more detailed presentation of the salient features of the theory; (3) an examination of criticisms that point to the theory's limitations and problems; and (4) an indication of the theory's strengths. This structure suggests that we accept several moral theories. We are pluralists in that we accept as legitimate various *aspects* of several different theories advanced in the history of ethics.[1] However, we reject the view that all the leading principles in the major moral theories can be rendered coherent (they cannot), as well as the view that the major theories offer equally tenable moral systems (they do not).

Criteria for Theory Construction

We begin with eight conditions of adequacy for an ethical theory. These proposals for theory construction set forth exemplary conditions for theories, but not so exemplary that a theory could not satisfy them. That all available theories only partially satisfy the demands in these conditions is not of concern here. The objective is to provide a basis from which to assess the defects and the strengths of theories. Satisfaction of these conditions protects a theory from criticism as a mere list of disconnected norms generated from our pretheoretic beliefs.The same general criteria of success in a moral theory can be used for any type of theory (for example, a scientific theory or a political theory). The eight conditions that follow express these criteria.[2]

1. Clarity. First, a theory should be as clear as possible, as a whole and in its parts. Although we can expect only as much precision of language as is appropriate, more obscurity and vagueness exists in the literature of ethical theory and biomedical ethics than is necessary or justified by the subject matter.

2. Coherence. Second, an ethical theory should be internally coherent. There should be neither conceptual inconsistencies (for example, "strong medical paternalism is justified only by consent of the patient") nor contradictory statements (for example, "to be virtuous is a moral obligation, but virtuous conduct

is not obligatory"). Ralph Waldo Emerson dismissed a foolish consistency as "the hobgoblin of little minds, adored by little statesmen and philosophers and divines." However, consistency is not a *sufficient* condition of a good theory, only a *necessary* condition. If an account has implications that are incoherent with other established parts of the account, some aspect of the theory needs to be changed in a way that does not produce further incoherence. Following the analysis in Chapter 1, a major goal of a theory is to bring into coherence all its various normative elements (principles, rights, considered judgments, etc.).

3. Completeness and Comprehensiveness. A theory should be as complete and comprehensive as possible. A theory would be entirely comprehensive if it included all moral values. Any theory that includes fewer moral values will be somewhere on a continuum from partially complete to void of important values. Although the principles presented in this book under the headings of respect for autonomy, nonmaleficence, beneficence, and justice are far from a complete system for general normative ethics, they do, when specified, provide a sufficiently comprehensive general framework for *biomedical* ethics. We do not need additional principles such as promise keeping, avoiding killing, keeping contracts, and the like. However, we draw on our principles to help justify rules of promise keeping, truthfulness, privacy, and confidentiality, among others (see esp. Chapter 7), and these norms increase the system's comprehensiveness by specifying commitments in the fundamental principles, as *specification* is defined in Chapter 1.

4. Simplicity. If a theory with a few basic norms generates sufficient moral content, then that theory is preferable to a theory with more norms but no additional content. A theory should have no more norms than are necessary, and no more than people can use without confusion. However, morality is complicated, and any comprehensive moral theory will be immensely complex. We can demand only as much simplicity in a moral theory as its subject matter permits.

5. Explanatory Power. A theory has explanatory power when it provides enough insight to help us understand the moral life: its purpose, its objective or subjective status, how rights are related to obligations, and the like.

6. Justificatory Power. A theory should also give us grounds for *justified* belief, not a reformulation of beliefs we already possess. For example, the distinction between acts and omissions underlies many critical beliefs in biomedical ethics, such as the belief that killing is impermissible and allowing to die permissible. But a moral theory would be impoverished if it only expressed this distinction without determining whether the distinction justifiably grounds those beliefs. A good theory also should have the power to criticize defective beliefs, no matter how widely accepted those beliefs may be.

7. Output Power. A theory has output power when it produces judgments that were not in the original data base of particular and general considered judgments on which the theory was constructed. If a theory did no more than repeat the list of judgments thought to be sound prior to the construction of the theory, nothing would have been accomplished. For example, if the parts of a theory pertaining to obligations of beneficence do not yield new judgments about role obligations of care in medicine beyond those assumed in constructing the theory, this failure of output suggests that the theory is purely a classification scheme. A theory, then, must generate more than a list of the axioms present in pretheoretic belief.

8. Practicability. A proposed moral theory is unacceptable if its requirements are so demanding that they probably cannot be satisfied or could be satisfied by only a few extraordinary persons or communities. A moral theory that presents utopian ideals, paltry expectations, or unfeasible recommendations fails the criterion of practicability. For example, if a theory proposed such high requirements for personal autonomy (see Chapter 3) or such lofty standards of social justice (see Chapter 6) that, realistically, no person could be autonomous and no society could be just, the proposed theory would be deeply defective.

Other general criteria could be formulated, but the eight sketched above are the most important for our purposes. A theory can receive a high score on the basis of one criterion and a low score on the basis of another. For example, early in this chapter utilitarianism is depicted as an internally coherent, simple, and comprehensive theory with exceptional output power, yet it is not coherent with some of our vital considered judgments, especially with certain judgments about justice, human rights, and the importance of personal projects. By contrast, Kantian theories are consistent with many of our considered judgments, but their clarity, simplicity, and output power are limited.

A contested and appropriately criticized moral theory may nonetheless be defensible in light of the criteria we have proposed. Although we currently have no perfect or even best moral theory, several good theories are available.

Utilitarianism: Consequence-Based Theory

Consequentialism is a label affixed to theories holding that actions are right or wrong according to the balance of their good and bad consequences. The right act in any circumstance is the one that produces the best overall result, as determined from an impersonal perspective that gives equal weight to the interests of each affected party. The most prominent consequence-based theory, utilitarianism, accepts one and only one basic principle of ethics: the principle of utility. This principle asserts that we ought always to produce the maximal balance of positive value over disvalue (or the least possible disvalue, if only undesirable results can be achieved). The classical origins of this theory are

found in the writings of Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873).

At first sight, utilitarianism seems entirely compelling. Who would deny that evil should be minimized and positive value increased? Moreover, utilitarians offer many examples from everyday life to show that the theory is practicable and that we all engage in a utilitarian method of calculating what should be done by balancing goals and resources and considering the needs of everyone affected. Examples include designing a family budget to meet the family's needs and creating a new public park in a wilderness region. Utilitarians maintain that their theory renders explicit and systematic what is already implicit in everyday deliberation and justification.

The Concept of Utility

Although utilitarians share the conviction that human actions should be morally assessed in terms of their production of maximal value, they disagree concerning which values are most important. Many utilitarians maintain that we ought to produce *agent-neutral or intrinsic* goods—that is, the goods every rational person values.[3] These goods are valuable in themselves, without reference to their further consequences or to the particular values held by individuals.

Bentham and Mill are *hedonistic* utilitarians because they conceive utility entirely in terms of happiness or pleasure, two broad terms they treat as synonymous.[4] They appreciate that many human actions are apparently not performed for the sake of happiness. For example, when highly motivated professionals, such as research scientists, work themselves to the point of exhaustion in search of new knowledge, they often do not appear to be seeking pleasure or personal happiness. Mill proposes that such persons are initially motivated by success or money, both of which promise happiness. Along the way, either the pursuit of knowledge provides pleasure, or such persons never stop associating their hard work with the success or money they hope to gain.

However, many recent utilitarian philosophers have argued that values other than happiness have intrinsic worth. Some list friendship, knowledge, health, and beauty among these intrinsic values, whereas others list personal autonomy, achievement and success, understanding, enjoyment, and deep personal relationships.[5] Even when their lists differ, these utilitarians concur that the greatest good should be assessed in terms of the total intrinsic value produced by an action. Still other utilitarians say that the concept of utility does not refer to intrinsic goods, but to an individual's preferences.

A Case of Risk and Truthfulness

To sketch the major themes of each theory, each section in this chapter devoted to a theory explicates how its proponents might approach the same case, which

centers on a five-year-old girl who has progressive renal failure and is not doing well on chronic renal dialysis. The medical staff is considering a renal transplant, but its effectiveness is "questionable" in her case. Nevertheless, a "clear possibility" exists that the transplanted kidney will not be affected by the disease process. The parents concur with the plan to try a transplant, but an additional obstacle emerges: The tissue typing indicates that it would be difficult to find a match for the girl. The staff excludes her two siblings, ages two and four, as too young to be donors. The mother is not histocompatible, but the father is compatible and has "anatomically favorable circulation for transplantation."

Meeting alone with the father, the nephrologist gives him the results and indicates that the prognosis for his daughter is "quite uncertain." After reflection, the father decides that he does not wish to donate a kidney to his daughter. His several reasons include his fear of the surgery, his lack of courage, the uncertain prognosis even with a transplant, the slight prospect of a cadaver kidney, and the suffering his daughter has already sustained. The father then asks the physician "to tell everyone else in the family that he is not histocompatible." He is afraid that if family members know the truth, they will accuse him of intentionally allowing his daughter to die. He maintains that truth-telling would have the effect of "wrecking the family." The physician is uncomfortable with this request, but after further discussion he agrees to tell the man's wife that "for medical reasons the father should not donate a kidney."[6]

Utilitarians evaluate this case in terms of the consequences of the different courses of action open to the father and the physician. The goal is to find the single greatest good by balancing the interests of all affected persons. This evaluation depends on judgments about probable outcomes. Whether the father ought to donate his kidney depends on the probability of successful transplantation as well as the risks and other costs to him (and indirectly to other dependent members of the family). The probability of success is not high. The effectiveness is questionable and the prognosis uncertain, although a possibility exists that a transplanted kidney would not undergo the same disease process, and there is a slight possibility that a cadaver kidney could be obtained.

The girl will probably die without a transplant from either a cadaveric or a living source, but the transplant also offers only a chance of survival. The risk of death to the father from anesthesia in the kidney removal is 1 in 10,000–15,000; it is difficult to put an estimate on other possible long-term health effects. Nevertheless, with a sufficiently high probability of success and a sufficiently low probability of harm, many utilitarians would hold that the father or anyone else similarly situated is *obligated* to undertake what many would consider a heroic act that *surpasses* obligation. On a certain balance of probable benefits and risks, an uncompromising utilitarian would suggest tissue typing the patient's two siblings and then removing a kidney from one if there

were a good match and parental approval. However, utilitarians differ among themselves in these various judgments because of different theories of value and different assessments of probable outcomes.

Probabilistic judgments would likewise play a role in the physician's utilitarian calculation of the right action in response to the father's request to camouflage why he will not donate a kidney. Primary questions include whether a full disclosure would actually wreck the family, whether lying to the family would have serious negative effects, and whether the father would subsequently experience serious guilt from his refusal to donate, thereby jeopardizing relations within the family. Studies indicate that families caring for chronically ill children break up at a higher rate than other families, and perhaps this family is already beyond repair. The utilitarian holds that the physician is obligated to consider the whole range of facts and possible consequences in light of the best available information about their probability and magnitude.

So far we have taken primarily the perspective of a utilitarian who focuses on *particular acts*. Other utilitarians focus on the relevant *principles and rules* of parental obligation and professional practice that, over time, maximize overall welfare. We turn now to this distinction between different types of utilitarian theory.

Act and Rule Utilitarianism

The principle of utility is the ultimate standard of rightness and wrongness for all utilitarians. Controversy has arisen, however, over whether this principle pertains to particular acts in particular circumstances or instead to general rules that determine which acts are right and wrong. Whereas the rule utilitarian considers the consequences of adopting rules, the act utilitarian skips the level of rules and justifies actions by appealing directly to the principle of utility, as the following chart indicates:

Rule Utilitarianism	*Act Utilitarianism*
Principle of Utility	Principle of Utility
↑	
Moral Rules	↑
↑	
Particular Judgments	Particular Judgments

The act utilitarian asks, "What good and bad consequences will result from this action in this circumstance?," not "What good and bad consequences will result from this sort of action in such circumstances?" The act utilitarian sees moral rules as somewhat useful in guiding human actions, but also as expendable if they do not promote utility in a particular context. For the rule utilitarian, by contrast, an act's conformity to a justified rule (that is, a rule justified by utility) makes the act right, and the rule is not expendable in a particular context, even if following the rule in that context does not maximize utility.

Physician Worthington Hooker, a prominent nineteenth-century figure in academic medicine and medical ethics, was a rule utilitarian who attended to rules of truth-telling in medicine as follows:

> The good, which may be done by deception in a *few* cases, is almost as nothing, compared with the evil which it does in *many*, when the prospect of its doing good was just as promising as it was in those in which it succeeded. And when we add to this the evil which would result from a *general* adoption of a system of deception, the importance of a strict adherence to the truth in our intercourse with the sick, even on the ground of expediency, becomes incalculably great.

Hooker agreed that a patient's health is sometimes maximally advanced through deception, but he argued that a widespread use of deception in medicine will have an increasingly negative effect over time and will eventually cause more harm than good. He therefore defended the rule-utilitarian conclusion that deception should be prohibited in medicine.[7]

Act utilitarians, by contrast, argue that observing a rule such as truth-telling does not always maximize the general good, and that the rule is properly understood as a rough guideline. They regard rule utilitarians as unfaithful to the fundamental demand of the principle of utility: Maximize value. In some circumstances, they argue, abiding by a generally beneficial rule will not prove most beneficial to the persons affected by the action, even in the long run. Why, then, should a rule be obeyed if obedience will not maximize value? According to a contemporary act utilitarian, J. J. C. Smart, a third possibility exists between never adopting any rules and always obeying rules; namely, *sometimes* obeying rules.[8] From this perspective, physicians do not and should not always tell the truth to their patients or their families, just as the physician uses misleading language to protect the father in the above case. Sometimes physicians even must lie to give hope. They do so justifiably if it is better for the patients and for all concerned and if their acts do not undermine general conformity to moral rules. According to Smart, selective obedience does not erode either moral rules or general respect for morality. Rules, then, are stabilizing but nonbinding guides in the moral life.

Because of the benefits to society of the general observance of moral rules, the rule utilitarian does not abandon them even in difficult situations (although the rule utilitarian may accept rules only as statements of prima facie obligation). Abandonment threatens the integrity and existence of both the individual rules and the whole system of rules.[9] The act utilitarian's reply is that although promises usually should be kept in order to maintain trust, this consideration should be set aside in cases in which overall good would be produced by breaking the promise. The act utilitarian might also argue that making exceptions to accepted rules is consistent with ordinary moral beliefs, because we often make

exceptions to rules without acting wrongly. The act utilitarian also contends in some cases when breaking rules clashes with our considered moral convictions, we need to revise our ordinary convictions rather than discard act utilitarianism.

An example of the act utilitarian's point appears in a comment by former Colorado governor Richard Lamm, who once observed that in light of increasing financial costs of medical care the terminally ill have "a duty to die and get out of the way with all of our machines and artificial hearts and everything else." This statement clearly conflicts with ordinary morality, and there was an outcry of indignation and shock that a public official would brush aside considered moral rules that protect our rights. Lamm chose an unfortunate word when he stated that the terminally ill have a "duty" to die. But in context he was giving an act-utilitarian answer to what he correctly referred to as an "ethical question." His point was that we cannot continue public funding for medical technology without assessing costs and trade-offs, even if we must subsequently revise our traditional views and let some people die because a technology is not funded. The act utilitarian believes that many other questions posed by technological developments likewise cannot be handled by traditional moral rules.

An Absolute Principle with Derivative Contingent Rules

From the utilitarian's perspective only the principle of utility is absolute. No derivative rule is absolute, and no rule is unrevisable. Even rules against killing in medicine may be overturned or substantially revised. For example, we will have occasion later to discuss current debates in biomedical ethics regarding whether seriously suffering patients should, at their request, be killed rather than "allowed to die," although such acts would revise traditional beliefs in medicine. The rule utilitarian argues that we should support rules permitting killing if and only if those rules would produce the most favorable consequences. Likewise, there should be rules against killing if and only if those rules would maximize good consequences. The utilitarian views euthanasia as a delicate matter of balancing risks and interests, whether in public policy or in particular judgments.

Imagine that a physician has a patient who requests to be killed, and the best utilitarian outcome would result from killing the patient. But suppose the physician cannot bring himself or herself to perform the act. Here the utilitarian will judge that the physician has not done the right thing, but may, in addition, note that good social consequences flow to society by having physicians who care so deeply about not causing harm to patients. Utilitarians often point out that we do not presently permit physicians to kill patients because of the adverse social consequences that we believe would be produced for those directly and indirectly affected. But if, under a different set of social conditions, legalization of mercy killing would maximize overall social welfare, the utilitarian sees no reason to prohibit such killing. Utilitarians for this reason regard their theory as responsive in constructive ways to changing social conditions.

A Critical Evaluation of Utilitarianism

Several problems suggest that utilitarianism is not a fully adequate moral theory.

Problems with immoral preferences and actions. Problems arise for utilitarians who appeal to preferences when individuals have what our considered judgments tell us are morally unacceptable preferences. For example, if a research investigator derived supreme satisfaction from inflicting pain on animals or on human subjects in experiments, we would condemn this preference and would seek to prevent it from being actualized. Utilitarianism based on subjective preferences is a defensible theory only if a range of acceptable preferences can be formulated, where "acceptability" is determined independently of the preferences of agents. This task seems inconsistent with a pure preference approach, but will utilitarianism be destroyed by a second level that delineates what counts as an acceptable preference?

There is an additional problem of immoral actions. Suppose the only way to achieve the maximal utilitarian outcome is to perform an immoral act (as judged by the standards of the common morality). For example, suppose a war can be ended only by using extremely painful methods of torturing captured children who were told by their soldier fathers not to reveal their location. Utilitarianism seems to say not only that you are permitted to torture the children, but that you are morally required to do so. Yet this requirement seems blatantly immoral. Thus, utilitarianism seems to permit apparently immoral actions without giving sufficient reasons for us to abandon our reigning views.

Does utilitarianism demand too much? Many forms of utilitarianism also seem to demand too much in the moral life, because the principle of utility is a *maximizing* principle. Utilitarians have a difficult time maintaining a crucial distinction between (1) *morally obligatory actions*, and (2) *supererogatory actions* (those above the call of moral obligation and performed for the sake of personal ideals). This objection has been registered by Alan Donagan, who describes situations in which utilitarians are committed by their theory to regard an action as obligatory against our firm moral conviction that the action is ideal and praiseworthy rather than obligatory.[10]

Donagan would regard suicides by the frail elderly and persons with severe disabilities who are no longer of use to society as examples of an act that could never rightly be considered obligatory, irrespective of its consequences. Heroic

donation of bodily parts such as kidneys and even hearts to save another person's life is another example. If utilitarianism makes such actions obligatory, then it is a defective theory. Donagan argues, and we agree, that all utilitarians face these problems, because none can rule out the ever-present possibility that what is today praiseworthy (but optional) will, through altered social circumstances, become obligatory by utilitarian standards. At the same time, we should recognize that utilitarians are sometimes right in arguing that ordinary morality is too weak or vague in its demands and should be upgraded by more demanding requirements.[11] Furthermore, in a changing social situation, our considered judgments may themselves undergo alteration.

Bernard Williams and John Mackie offer extensions of the thesis that utilitarianism demands too much. Williams argues that utilitarianism abrades personal integrity by making persons as morally responsible for consequences that they fail to prevent as for those outcomes they directly cause, even when the consequences are not of their doing. Mackie similarly argues that a utilitarian "test of right actions" is so distant from our moral experience as to be "the ethics of fantasy," because it demands that people strip themselves of many goals and relationships they value in life in order to maximize outcomes for others. From this perspective, the utilitarian demands that we act like saints who are without personal interests and goals.[12] These criticisms suggest that utilitarianism fails the test of *practicability* presented at the beginning of this chapter.

Problems of unjust distribution. A third problem is that utilitarianism in principle permits the interests of the majority to override the rights of minorities, and cannot adequately disavow unjust social distributions. The charge is that utilitarians assign no independent weight to justice and are indifferent to unjust distributions, instead insisting that value be distributed by net aggregate satisfaction. If an already prosperous group of persons could have more value added to their lives than the value that could be added to the lives of the indigent in society, the utilitarian must recommend that the added value go to the prosperous group.

An example of problematic (although not necessarily unjust) distribution appears in the following case. Two researchers wanted to determine the most cost-effective way to control hypertension in the American population. As they developed their research, they discovered that it is more cost-effective to target patients already being treated for hypertension than to identify new cases of hypertension among persons without regular access to medical care: younger men, older women, and patients with exceptionally high blood pressure. And they concluded that "a community with limited resources would probably do better to concentrate its efforts on improving adherence of known hypertensives, even at a sacrifice in terms of the numbers screened." If accepted by the government, this recommendation would exclude the poorest sector, which has the most pressing need for medical attention, from the benefits of high blood pressure education and management.

The investigators were concerned because of the apparent injustice in excluding the poor and minorities by a public health endeavor aimed at the economically advantaged sector of society. Yet their statistics were compelling. No matter how carefully planned the efforts, nothing worked efficiently (that is, nothing produced utilitarian results) except programs directed at known hypertensives already in contact with physicians. The investigators therefore recommended what they explicitly referred to as a utilitarian allocation.[13]

A Constructive Evaluation of Utilitarianism

Despite these criticisms, utilitarianism has many strengths, two of which we appropriate in later chapters. The first is the acceptance of a role for the principle of utility in the formation of public policy. The utilitarian's requirements of an objective assessment of everyone's interests and of an impartial choice to maximize good outcomes for all affected parties are acceptable norms of public policy. Second, when we formulate principles of beneficence in Chapter 5, utility plays an important role. Although we have characterized utilitarianism as primarily a *consequence*-based theory, it is also *beneficence*-based. That is, the theory sees morality primarily in terms of the goal of promoting welfare.

A theory with a principle of beneficence balanced by other principles should eliminate all the problems with an unqualified use of the principle of utility that we encountered in the criticisms offered in the preceding section. This point holds even if beneficence is developed primarily in terms of producing good consequences. As political economist Amartya Sen notes, "Consequentialist reasoning may be fruitfully used even when consequentialism as such is not accepted. To ignore consequences is to leave an ethical story half told."[14]

A *strict* or *pure* utilitarianism also has strengths, as we can see by reconsidering the objection that utilitarianism is overdemanding. Utilitarianism often demands more than the rules of the common morality do, but this apparent weakness is also a hidden strength. For example, ordinary morality demands that we not override the rights of individuals to maximize social consequences. But if we can more widely and more effectively protect almost everyone's interests by overriding some property and autonomy rights, then it is far from clear that this course of action would be wrong merely because it contravenes ordinary morality and pursues the goal of social utility. In many circumstances the utilitarian makes a compelling case in advising us to rely less on everyday convictions and more on judgments of overall benefit.

Kantianism: Obligation-Based Theory

A second type of theory denies much that utilitarian theories affirm. Often called *deontological* (i.e. a theory that some features of actions other than or in addition to consequences make actions right or wrong), this type is now increasingly called *Kantian*, because the ethical thought of Immanuel Kant (1724–1804), has shaped many of its formulations.

Consider how a Kantian might approach the above-mentioned case of the five-year-old in need of a kidney. A Kantian would first insist that we rest our moral judgments on reasons that can be generalized for others who are similarly situated. If the father has no generalizable moral obligation, no basis is available for moral criticism of him. The strict Kantian takes this point to be a rigid demand. If the father chooses to donate out of affection, compassion, or concern for his dying daughter, his act would actually lack moral worth, because it would not be based on a recognition of generalizable obligation. It would also not be legitimate to use one of the girl's younger siblings as a source of a kidney, because that recourse would involve using persons entirely as means to others' ends. This same principle would also exclude coercing the father to donate against his will.

Regarding the physician's options after the father requests deception of the family, a strict Kantian views lying as an act that cannot without contradiction be universalized as a norm of conduct. Thus, the physician should not lie to the man's wife or to other members of the family, even if the lie would function to salvage the family (a consequentialist appeal). Even if the physician's statement is not, strictly speaking, a lie, he intentionally used this formulation to conceal relevant facts from the wife, an act Kantians typically view as morally unacceptable.

A Kantian will also consider whether the rule of confidentiality has independent moral weight, whether the tests the father underwent with the nephrologist established a relationship of confidentiality, and whether the rule of confidentiality protects the information about the father's histocompatibility and his reasons for not donating. Even without considering possible effects on the family, the Kantian seems destined to face a difficult conflict of obligations: truthfulness in conflict with confidentiality. But before we can address a possible Kantian strategy for resolving this conflict, we need to understand more about Kantian theory.

Obligation from Categorical Rules

In an attempt to combat skeptical challenges to ethics, Kant argued that morality is grounded in pure reason, not in tradition, intuition, conscience, emotion, or attitudes such as sympathy. Kant saw human beings as creatures with rational powers to resist desire, the freedom to do so, and the capacity to act by rational considerations. He held that the moral worth of an individual's action depends exclusively on the moral acceptability of the rule (or "maxim") on which the person acts. As Kant puts it, moral obligation depends on the rule that determines the individual's will. An action possesses moral worth only if performed by an agent with a good will, which entails that a morally valid reason justify the action.[15]

For Kant, one must act not only *in accordance with* but *for the sake of* obligation. That is, to have moral worth, a person's motive for acting must come from a recognition that he or she intends what is morally required. For example, if an employer discloses a health hazard to an employee only because the employer fears a lawsuit, and not because of the importance of truth-telling or concern about the employee's health, then the employer has done the right thing but deserves no moral credit for the action. If agents do what is morally right simply because they are scared, because they derive pleasure from doing that kind of act, or because they are selfish, they lack the requisite good will that derives from acting for the sake of obligation.

Imagine a man who desperately needs money and knows that he will not be able to borrow it unless he promises repayment in a definite time, but who also knows that he will not be able to repay it within this period. He decides to make a promise that he knows he will break. Kant asks us to examine the man's reason, what Kant calls the maxim of the action: "When I think myself in want of money, I will borrow money and promise to pay it back, although I know that I cannot do so." This maxim, Kant says, cannot pass a test that he calls the *categorical imperative*. This imperative tells us what must be done irrespective of our desires. It requires unconditional conformity by all rational beings. In its major formulation, Kant states the categorical imperative as follows: "I ought never to act except in such a way that I can also will that my maxim become a universal law." Kant says that all particular imperatives of obligation (all "ought" statements that morally obligate) are justified by this one principle.

The categorical imperative, then, is a canon of the acceptability of moral rules—that is, a criterion for judging the acceptability of the maxims that direct actions.[16] This imperative adds nothing to a maxim's content. Rather, it determines which maxims are objective and valid. The categorical imperative functions by testing what Kant calls the consistency of maxims: A maxim must be capable of being conceived and willed without contradiction. When we examine the maxim of the person who deceitfully promises, we discover, according to Kant, that this maxim is incapable of being conceived and willed without contradiction. It is inconsistent with what it presupposes. The maxim would make the purpose of promising impossible, because no one would believe promises. Many examples from everyday life illustrate this thesis. For instance,

maxims of lying are inconsistent with the practices of truth-telling they presuppose, and maxims permitting cheating on tests are inconsistent with the practices of honesty they presuppose.

Kant appears to have more than one categorical imperative, because his several formulations are not equivalent. His second formulation is at least as influential as the first: "One must act to treat every person as an end and never as a means only."[17] It has often been said that this principle categorically requires that we should never treat another as a means to our ends, but this interpretation misrepresents his views. He argues only that we must not treat another *exclusively* as a means to our ends. When secretaries type manuscripts and human research subjects volunteer to test new drugs, they are treated as a means to others' ends, but they have a choice in the matter and retain control over their lives. Kant does not prohibit such uses of consenting persons. He insists only that they be treated with the respect and moral dignity to which every person is entitled.

Autonomy and Heteronomy

In contemporary biomedical ethics the word *autonomy* typically refers to what makes a life one's own; viz. that it is shaped by personal preferences and choices. This conception of autonomy is emphatically not Kant's. A person has "autonomy of the will" for Kant if and only if the person knowingly acts in accordance with the universally valid moral principles that pass the requirements of the categorical imperative. He contrasts this *moral* autonomy with "heteronomy," which is any controlling influence over the will other than motivation by moral principles.[18] If, for example, a person acts from passion, ambition, or self-interest, the person acts heteronomously, not from a rational will that chooses autonomously. Kant thus regards acting from desire, fear, impulse, personal projects, and habit as no less heteronomous than actions manipulated or coerced by others.

To say that an individual must "accept" a moral principle in order to be autonomous does not mean that the principle is subjective or that each individual must create (author or originate) his or her moral principles. Kant only requires that each individual *will the acceptance* of moral principles. By contrast, Kant's theory is exclusively one of the moral self-legislation of objective rules. If a person freely accepts objective moral principles, that person is a law-giver unto himself or herself. The importance of this account for Kant also extends beyond the *nature* of autonomy to its *value*. "The principle of autonomy," he holds, is "the sole principle of morals," and autonomy alone gives people respect, value, and proper motivation. A person's dignity—indeed, "sublimity"—comes from being morally autonomous.[19]

Contemporary Kantian Ethics

Several writers in contemporary ethical theory have accepted and developed a Kantian account, broadly construed. A straightforward example is *The Theory of Morality* by Alan Donagan. He seeks the "philosophical core" of the morality expressed in the Hebrew-Christian tradition, now interpreted in secular rather than religious terms. Donagan's philosophical elaboration of this point of view relies heavily on Kant's theory of persons as ends in themselves, especially the imperative that one must treat humanity as an end and never as a means only. Donagan expresses the fundamental principle of the Hebrew-Christian tradition as a Kantian principle grounded in rationality: "It is impermissible not to respect every human being, oneself or any other, as a rational creature."[20] Donagan believes that all other moral rules rely upon this fundamental principle and that Kant's theory captures the rational basis of these rules.

A second theory has encouraged the use of Kantian insights in contemporary ethics. John Rawls, whose theory of reflective equilibrium was examined in Chapter 1, challenges utilitarian theories while attempting to develop Kantian themes of reason, autonomy, equality, and opposition to utilitarianism. For example, Rawls argues that vital moral considerations, such as individual rights and the just distribution of goods among individuals, depend less on social factors, such as individual happiness and majority interests, than on Kantian conceptions of individual worth, self-respect, and autonomy.[21]

For Rawls, a social arrangement is a communal effort to advance the good of all. Inequalities of birth, natural endowment, and historical circumstance are undeserved, and persons in a cooperative society should make more equal the situation of persons disadvantaged through no fault of their own. Those who are naturally endowed with more advantageous properties by luck of the draw do not deserve their advantageous properties, and hence a just society would seek in its scheme of justice to overcome advantages stemming from the accidents of biology and history. Rawls uses a hypothetical social contract, in which valid principles are those to which we would all agree if we could freely consider the social situation from a standpoint he calls the "original position," in which individuals are equally ignorant of the particular characteristics and advantages they do or will possess. They know that they live together in a cooperative venture, but they are blinded to their individual desires, interests, and objectives. In Kant's terms, they are purely rational agents behind what Rawls calls a "veil of ignorance."[22]

Rawls aligns his original position with the Kantian theory of autonomy. Individuals give themselves the law from the perspective of rationality alone. Autonomy is moral self-legislation through a structure of reason and will that is common to all rational agents. Persons are autonomous in the original position

because they choose and give to themselves the moral law out of their nature as rational, independent, and mutually disinterested persons. While treating Kantian autonomy, Rawls considers Henry Sidgwick's objection to Kant, that the principles of the scoundrel and the principles of the saint could both be accepted autonomously.[23] Rawls appropriately argues that this objection springs from a misunderstanding of Kantian theory. Although a free self *could* choose as a scoundrel would, this choice would be inconsistent with the choices that rational beings expressing their nature as such would make. For Rawls, any philosophy in which the right to individual autonomy legitimately outweighs the dictates of objective moral principles is unacceptable. Even courageous and conscientious actions do not merit respect unless they accord with objective moral principles. If society restricts conscientious actions that violate valid public principles, "no violation of our [moral] autonomy" exists, because these acts are not morally autonomous—no matter how freely and conscientiously chosen.[24]

In his recent writings, Rawls has stressed that his work presents a political conception of justice, rather than a comprehensive moral theory. That is, his account is "a moral conception worked out for a specific subject, namely, the basic structure of a constitutional democratic regime." As such, it does not presuppose a comprehensive moral doctrine such as Kant's. Rawls maintains that his theory is Kantian by "analogy not identity." He points to several Kantian perspectives with which he identifies, including the priority of the right over the good and persons as free, equal, and capable of autonomy.[25] The upshot seems to be that Rawls is expressing Kantian themes without making a full commitment to a Kantian or deontological moral theory. The same can be said of many other contemporary Kantians.

A Critical Evaluation of Kantianism

Like utilitarianism, Kantian theory fails to provide a full and adequate theory of the moral life, for reasons we shall now discuss.

The problem of conflicting obligations. Kant has a problem with conflicting obligations. Suppose we have promised to take our children on a long-anticipated trip, but now find that if we do so, we cannot assist our sick mother in the hospital. This conflict is generated between a rule of promise-keeping and a rule of assistance, perhaps based on a debt of gratitude. The conflict sometimes arises from a single moral rule rather than from two different rules in conflict—as, for example, when one has made two promises that now come into conflict, although one could not have anticipated the conflict at the time one made the promises.

Because he makes all moral rules absolute, Kant often seems to say that we

are obligated to do the impossible and perform both actions. We cannot at the same time both take our children on a trip and help our mother in the hospital; yet Kant seems to require both. Any ethical theory that leads to this conclusion is incoherent, yet no clear path exists out of Kant's absolutistic framework. If even as many as two absolute rules exist, they will conflict on some occasions. Either we must accept a system with only one absolute, or we must give up absolutes altogether unless their meaning and scope can be specified to avoid conflict. (Our solution to this problem is found on pp. 104–106 below.)

Overemphasizing law, underemphasizing relationships. Kant's arguments concentrate on lawful obligations, and recent Kantian theories, such as Rawls's, feature a contractual basis for obligations. But whether freedom, choice, equality, contract, law, and other staples of Kantianism deserve to occupy such a central position in a moral theory is questionable. (They are, we can agree, central ingredients in legal and political theories.) These visions of the moral life fail to capture much in personal relationships, which generate various responsibilities. Among friends and family we rarely think or act in terms of law, contract, or absolute rules. This suggests that Kant's theory (as with utilitarianism) is better suited for relationships among strangers than for relationships among friends or other intimates. Parents, for example, do not see responsibilities to their children in terms of contracts, but in terms of care, needs, sustenance, and loving attachment. Only if all forms of moral relationship—and our moral sentiments, motivations, and virtues—could be reduced to a law-governed exchange would Kantian theory be defensible.

The limitations of the categorical imperative. Many immoral actions cannot be pronounced "contradictory" as easily as Kant's tidy examples suggest, and Kant's categorical imperative is both obscure and difficult to render functional in the moral life. Few philosophers would now hold, as Kant appears to, that universalizability is sufficient to determine the moral acceptability of rules, although many concur with him that universalizability is a necessary condition of ethical judgments, rules, and principles. As long as these questions hang over Kant's central principles, questions will persist about the theory's output, explanatory, and justificatory power.

Many arguments Kant adduces to explain the categorical imperative carry little conviction beyond those already convinced. His arguments are sometimes so unconvincing that he himself draws on a source outside the categorical imperative. For example, to argue against the moral acceptability of suicide, he maintains that suicide violates an obligation to God, because the suicide "leaves the post assigned him" as a "sentinel on earth" and "violates a holy trust." He notes that "as soon as we examine suicide from the standpoint of religion we immediately see it in its true light. . . . God is our owner; we are

His property.''[26] Kant's moral arguments from the categorical imperative, then, are unconvincing because his position often appears to rely on an external source, such as theology.

Abstractness without content. In Chapter 1 we mentioned Hegel's criticism that Kant's theory lacks the power to develop an ''immanent doctrine of obligations'' and obliterates all ''content and specification'' in favor of abstractness.[27] We agree that concepts such as ''rationality'' and ''humanity'' are too thin a basis for a determinate set of moral norms. Kant's relatively empty formalisms have little power to identify or assign specific obligations in almost any context of everyday morality, thereby raising questions about the theory's practicability. Both their abstractness and this impracticability provide reasons why method in ethics should start with considered judgments and then to specify principles and test moral claims in light of coherence.

A Constructive Evaluation of Kantianism

Kant held that any person who judges that X is morally required in one circumstance is thereby committed to the view that X is morally required in any relevantly similar circumstance. To be consistently committed to a moral system of rules and principles is a moral requirement that Kant analyzed with profound insight. The basic idea is that when a moral judgment is supported by good reasons, those reasons are good for any relevantly similar circumstance. As a point of consistency, this claim is undeniable, and it is far-reaching. Persons cannot act morally and make themselves privileged or exempt. Relevant differences exist across persons and groups, and there are valid exceptions to all general rules; but when persons are situated in relevantly similar ways, consistency requires that they use the same justifying reasons and treat persons in the same ways. If Kant had done nothing else than establish this point, he would have made a significant contribution to ethical theory.

Character Ethics: Virtue-Based Theory

Utilitarian and some Kantian theories attempt to shape various moral phenomena into integrated frameworks structured by a single dominant principle. Despite the attractiveness of their formulations, recent ethical theory has attended to some neglected moral phenomena, including character and virtue.[28] Whereas utilitarian and Kantian theories are principally expressed in the language of obligations and rights, with a focus on situations of choice, *character ethics* or *virtue ethics* emphasizes the agents who perform actions and make choices. Following the tradition of Plato and Aristotle, character ethics assigns virtuous character a preeminent position.

From this standpoint, the father's confession of a lack of courage in the case previously considered is relevant to an evaluation of him and his refusal to donate. But he has other reasons as well, some possibly involving self-deception. He points to his daughter's ''degree of suffering,'' which suggests that he believes she might be better off without a transplant. From this viewpoint, his motives are partially altruistic, not purely self-centered. However, this judgment of altruism seems feeble in light of his comment about failed courage and may not be sustainable because of the delicate balance of benefits and risks, burdens, and costs that the family and the daughter face. A defender of character ethics would still wonder whether the father was sufficiently compassionate and caring about her welfare. Failed courage seems to have overwhelmed compassion and fidelity in the father.

Several other judgments of virtue and character are possible in this case. We lack a full description of his wife, but the father apparently worried that she would be vindictive and unforgiving in accusing him of ''allowing his daughter to die.'' This belief underlies his request that the physician lie. The physician, we saw, focused on how the act of deception might compromise his integrity. Apparently he thought he could sidestep this problem, at least avoiding a severe compromise of integrity, by saying ''for medical reasons'' the father should not donate a kidney. However, the fact that he ''felt very uncomfortable'' about the man's request to conceal information indicates an ongoing concern about both truthfulness and moral integrity, two central virtues. Questions can also be raised about whether the physician deceived himself when he recognized and acted on a perilous distinction between a direct lie (for instance, ''he cannot donate because he is not histocompatible'') and deliberate, effective deception (''for medical reasons he should not donate'').

The Concept of Virtue

A *virtue* is a trait of character that is socially valued,[29] and a *moral virtue* is a trait that is morally valued. The fact that courage, for example, is a socially valued trait does not necessarily make courage a moral virtue. Moral courage occurs only if the context is a moral one, and it is not sufficient that social groups approve of a trait and regard it as moral. Moral reasons must support a claim or perception of moral virtue. Persons are sometimes disvalued in a community when they act virtuously, and robbers and pirates are sometimes admired in their communities for their meanness and churlishness. It is a mistake, then, to reduce moral virtue to whatever is socially approved.

Some have defined moral virtue as a disposition to act or a habit of acting in accordance with moral principles, obligations, or ideals.[30] The moral virtue of nonmalevolence, for example, is understood as the trait a person has of abstaining from causing harm to others when it would be wrong to harm them.

However, this definition unjustifiably makes virtues wholly derivative from principles and fails to capture the importance of motives. Virtue is intimately connected to characteristic motives. We care morally about how persons are motivated, and we care especially about their *characteristic* forms of motivation. Persons who are motivated by sympathy and personal affection, for example, meet our approval when others who act the same way but from different motives would not meet our approval. Properly motivated persons often do not merely follow rules; they also have a morally appropriate desire to act as they do.

Imagine that a person discharges an obligation *because* it is an obligation, but intensely dislikes being placed in a position in which the interests of others are overriding. This person does not love, feel friendly toward, or cherish others and respects their wishes only because obligation requires it. This person can nonetheless perform a morally right action and have a disposition to perform that action. All he or she needs is a disposition to follow rules and perform obligation. But if the motive is improper, a vital moral ingredient is missing; and if a person *characteristically* lacks this motivational structure, a necessary condition of virtuous character is absent. The act may be right and the actor without blame, but neither the person nor the act is virtuous. In short, it is possible to be disposed to do what is right, to intend to do it, and to do it, while also yearning to avoid doing it. Persons who characteristically perform *morally right actions* from such a motivational structure are not *morally virtuous* even if they always perform the morally right action.

Aristotle expressed an important (although underdeveloped) distinction between right action and proper motive, which he also analyzed in terms of the distinction between external performance and internal state. An action can be right without being virtuous, he said, but an action can be virtuous only if performed from the right state of mind of the person. Both right action and right motive should be present in a virtuous action: "The agent must . . . be in the right state when he does [the actions]. First, he must know [that he is doing virtuous actions]; second, he must decide on them, and decide on them for themselves; and third, he must also do them from a firm and unchanging state," including the right state of emotions and desires. "The just and temperate person is not the one who [merely] does these actions, but the one who also does them in the way in which just or temperate people do them."[31]

Our analysis of the virtues in terms of motivational structure needs supplementation in the light of Aristotle's observations. First, in addition to being properly *motivated* to action, a virtuous person often must experience appropriate *feelings*, such as sympathy and regret—even when the feelings are not motives and no action can result from the feelings. Second, many virtues have no clear link to either motives or feelings. Moral discernment and moral integrity—two of the primary virtues treated in Chapter 8 (where we return to these problems)—are typical examples. Here behavior and psychological properties other than motives and feelings are paramount.

A Special Place for the Virtues

Some writers in character ethics maintain that the language of obligation is *derivative* from moral circumstances in which persons display a lack of virtue in not performing certain actions. Accordingly, one who is disposed by character to have the right motives and desires is the basic model of the moral person.[32] This model is more important, they claim, than a model of action-from-obligation, because right motives and character tell us more about moral worth than do right actions.

This position is attractive, because we are often more concerned about the character and motives of persons than about the conformity of their acts to rules. When a friend performs an act of "friendship," we expect it not to be motivated entirely from a sense of obligation to us, but rather because the person has a desire to be friendly, feels friendly, wants to keep friends in good cheer, and values friendship. The friend who acts only from obligation lacks the virtue of friendliness, which is vital. Absent this virtue, the relationship lacks moral merit.[33]

Some writers in biomedical ethics have also argued that the attempt in obligation-oriented theories to replace the virtuous judgments of health care professionals with rules, codes, or procedures will not result in better decisions and actions. For example, rather than using rules and government regulations to protect subjects in research, some claim that the most reliable protection is the presence of an "informed conscientious, compassionate, responsible researcher."[34] The underlying view is that character is more important than conformity to rules and that virtues should be inculcated and cultivated over time through educational interactions, role models, and the like.

Gregory Pence contends that moral issues in medicine and health care should be discussed in the framework of virtues, because almost any health professional can successfully evade a system of rules. We should, he says, create a climate in which health professionals "desire not to abuse their subjects—a point harking back to our definition of the good person as one who has the right kind of desires."[35] This argument provides a significant reason for incorporating the virtues into biomedical ethics and in medical and nursing education, but it needs elaboration.

A morally good person with the right configuration of desires and motives is more likely than others to understand what should be done, more likely to attentively perform the acts that are required, and even more likely to form and act on moral ideals. A person we trust is one who has an ingrained motivation and desire to perform right actions. Not the rule follower, then, but the person disposed by character to be generous, caring, compassionate, sympathetic, fair,

and the like, is the one we will recommend, admire, praise, and hold up as a moral model.

If a virtuous person makes a mistake in judgment, thereby performing a morally wrong act, he or she would be less blameworthy than an habitual offender who performed the same act. The person's character informs our judgment of the individual and how we assess his or her actions. In his chronicle of life under the Nazi SS in the Jewish ghetto in Cracow, Poland, Thomas Keneally describes a physician faced with a grave dilemma: either inject cyanide into four immobile patients or abandon them to the SS, who were at that moment emptying the ghetto and had already proved that they would brutally kill all captives and patients. This physician, Keneally reports, "suffered painfully from a set of ethics as intimate to him as the organs of his own body."[36] Here is a person of the highest moral character and virtue, motivated to act rightly and even heroically, yet who at first had no idea what was the morally right action.

Ultimately, with uncertainty and reluctance, the physician elected active euthanasia without the consent or knowledge of the four doomed patients (using forty drops of hydrocyanic acid)—an act almost universally denounced by the canons of professional medical ethics. Even if one thinks that the physician's act was wrong and blameworthy, a judgment we reject, no one could reasonably make a judgment of blame or demerit directed at the physician's motives or character. Having already risked death by choosing to remain at his patients' beds in the hospital rather than take a prepared escape route, this physician is a moral hero who has displayed an extraordinary moral character.

Judgments of an agent's merit and praiseworthiness or demerit and blameworthiness are tied to the person's motives, not merely to the person's actions. To speak of a good, praiseworthy, or virtuous action is elliptical for our evaluation of the motive underlying the action—for example, the motive of benefiting another person.[37] However, in contrast to some radical forms of character ethics, the merit in an action is not in motive or character alone. The action must be appropriately gauged to bring about the desired result and must be morally justified in conformity with relevant principles and rules. For example, the physician who is appropriately motivated to help a patient but who acts inappropriately to bring about the desired result does not act in a praiseworthy manner.

The Compatibility of Virtues and Principles

Although the virtues do have a special place in the moral life, this fact is not sufficient evidence for an exclusive primary role, as if a virtue-based theory could replace or take precedence over obligation-based theories. The two kinds of theory have different emphases, but they are compatible and mutually reinforcing. As the case of the Cracow physician shows, persons of good moral character sometimes have trouble discerning what is right and may be the first to recognize that they need principles, rules, and ideals to determine right or good acts.

In circumstances of conflicting motivation from different virtues, we also need to ask questions about which action is right, best, or obligatory. One often cannot act virtuously unless one makes judgments about the best ways to manifest sympathy, desire, and the like.[38] Consider what a generous and tolerant person would do in a circumstance in which outrage or punishment is an appropriate response to someone's wrongdoing. It would be improper behavior to be generous or tolerant toward the wrongdoer. To see that normally appropriate responses are here wrong requires a balancing of conflicting values. Such judgments are based on general norms, not on the virtues alone. This suggests that the virtues need principles and rules to regulate and supplement them. As Aristotle suggests, ethics involves judgments like those in medicine: Principles guide us to actions, but we still need to assess a situation and formulate an appropriate response, and this assessment and response flow from character and training as much as from principles.

To defend the compatibility of virtues and principles is not to argue for a perfect correspondence. That is, one need not argue that every moral virtue has a corresponding moral principle of obligation. The proposal that there might be such a correspondence is displayed in schematic form in the following diagram (in which "exceptional standards" are moral ideals, as discussed in Chapter 8).[39]

	Action Guides	[correspond to]	*Virtue Standards*
Ordinary Standards	Principles or Rules of Obligation		Virtue Standards
Exceptional Standards	Ideals of Action		Ideals of Virtue

The following list illustrates the correspondence between some specific action guides and virtues.

Principles	*Corresponding Virtues*
Respect for Autonomy	Respectfulness
Nonmaleficence	Nonmalevolence
Beneficence	Benevolence
Justice	Justice or Fairness
Rules	*Corresponding Virtues*
Veracity	Truthfulness
Confidentiality	Confidentialness

Privacy	Respect for Privacy
Fidelity	Faithfulness
Ideals of Action	*Ideals of Virtue*
Exceptional Forgiveness	Exceptional Forgiveness
Exceptional Generosity	Exceptional Generosity
Exceptional Compassion	Exceptional Compassion
Exceptional Kindness	Exceptional Kindness

This list could be expanded to include many additional action-guides and virtues, but a systematic program of correspondence likely cannot be developed from this programatic idea. Many virtue-standards do not directly correspond to action-guides. No one-to-one correspondence exists, even if there is some form of relationship. For example, concern, compassion, caring, sympathy, courage, modesty, and patience do not correspond to norms of obligation. This problem is broader than an absence of one-to-one relationships. Many virtues seem to have no direct connection to norms of obligation, although they contribute to or even improve actions done from obligation. Typical examples are cautiousness, integrity (in the sense of consistently upholding and standing firm in one's values), cheerfulness, unpretentiousness, sincerity, appreciativeness, cooperativeness, and commitment.

A Critical Evaluation of Character Ethics

We can now investigate some limitations of virtue theories.

Morality in relations between strangers. Not all areas of the moral life can be forced into the language and the framework of virtue theory without a loss of vital moral protections. Character judgments will often play a less significant role than rights and procedures (such as committee review), especially when strangers meet. For example, when a patient first encounters a physician, the physician's conformity to rules or principles (and even explicit contracts backed by sanctions) may be essential for their subsequent relationship. This reliance on principles and rules does not presuppose an unacceptable form of distrust. A presumption of trust can be combined with a recognition that people who are typically trustworthy at least occasionally need guidance from principles and rules.

Virtue is not enough. The first criticism leads to a second. It is doubtful that character ethics can adequately *explain* and *justify* assertions of the rightness or wrongness of specific actions. It is unacceptable to claim that if persons display a virtuous character, their acts are therefore morally acceptable. People of good character who act virtuously can perform wrong actions. They may have incorrect information about likely consequences, make incorrect judgments, or fail to grasp what should be done. Defenders of character ethics cannot plausibly maintain that just and unjust actions consist only in what just and unjust persons do. We sometimes cannot even evaluate a motive as being appropriate unless we know that certain forms of conduct are obligatory, prohibited, or permissible.

The defender of a pure character ethics must say that the virtues themselves, not principles or rules, guide action. The strength of this account is in the strength of character of the virtuous person. In a virtuous person who is decisive and resolute, this character should prove to be as functional in guiding action as rules and principles. When confronted with the question, which actions should be performed, a virtue theory can answer, "those actions that an exemplary moral agent would perform."[40] Although we deeply respect this point (and develop it further in Chapter 8), it needs qualification. In many circumstances, principles and rules are essential to guide conduct.

A Constructive Evaluation of Character Ethics

A proponent of character ethics need not claim that analysis of the virtues subverts or discredits ethical principles and rules. It is enough to argue that ethical theory is more complete if the virtues are included and that moral motives deserve to be at center stage in a way some leading traditional theories have inadequately appreciated. When the feelings, concerns, and attitudes of others are the morally relevant matters, rules and principles are not as likely as human warmth and sensitivity to lead us to notice what should be done. Even a seldom noticed virtue, such as cheerfulness or tactfulness, can be far more significant than standard rules in some contexts. Furthermore, forms of loyalty, reliability, and commitment to other persons can, across time, be more integral to an adequate or full moral life than following principles or rules.

To look at acts without also looking at the moral appropriateness and desirability of feelings, attitudes, forms of sympathy, and the like is to miss a large area of the moral picture. We do not merely expect persons to act in certain ways. We also expect them to have certain emotions, certain forms of responsiveness, and a trustworthy character. Character ethics helps us introduce this subtlety in moral theory, as we will see in later chapters.

Liberal Individualism: Rights-Based Theory

Thus far we have primarily been using terms such as the following from moral discourse: *obligation, permissible action, virtue,* and *justification.* It may seem odd that we have not often used the language of rights, given their historical importance and their recent role in ethics and foreign policy. Statements of rights provide vital protections of life, liberty, expression, and property. They

protect against oppression, unequal treatment, intolerance, arbitrary invasion of privacy, and the like. Many philosophers and framers of political declarations therefore regard rights language as supplying the basic terminology for expressing the moral point of view.

An ethical analysis of the case of the five-year-old needing a transplant would, from this perspective, focus on the rights of all the parties, in an effort to determine the meaning and scope as well as the weight and strength of those rights. The father could be viewed as having rights of autonomy, privacy, and confidentiality that call for the protection of his bodily integrity and sphere of decisionmaking from interference by others. In addition, he has a right to information, which he apparently received, about the risks, benefits, and alternatives of living kidney donation.

The father's decision not to donate is within his rights, as long as it does not violate another's rights. No apparent grounds support a right to assistance that could permit anyone, including his daughter, to demand a kidney. However, there are some special rights to assistance, and it could be argued that the daughter has a right to receive a kidney from her father, on the basis of either parental obligations or medical need. But even if such a right exists, it would be sharply bounded. For example, it is implausible to suppose that such a right could be enforced against the girl's two siblings. Their right to noninterference, when the procedure is not for their direct benefit and carries risks, protects them against recruitment as sources for a kidney.

An analysis in terms of rights might also notice that the father exercises his rights of autonomy and privacy in allowing the physician to run some tests, and then seeks protection behind a right of confidentiality, which allows him to control further access to information generated in his relationship with the physician. The scope and limits of those rights and of competing rights need attention. For example, does the mother have a right to the information generated in the relationship between the father and the nephrologist, particularly information bearing on the fate of the daughter?

An analysis using rights would also consider whether the physician has a relevant right of conscience. For example, the physician might resist becoming an instrument of the father's desire to keep others from knowing why he is not donating a kidney. But even if the physician does have a right to protect his integrity, does this right outstrip or trump the rights of others? Can a physician justifiably say "I have a right of conscience" and use this trump to back out of a moral dilemma?

The Nature of Liberal Individualism

Rights theory will here be analyzed as liberal individualism, the conception that in a democratic society a certain space must be carved out within which

the individual is protected and allowed to pursue personal projects. Liberal individualism has, in recent years, challenged the reigning utilitarian and Kantian models. H. L. A. Hart has described this challenge as a switch from an "old faith that some form of utilitarianism . . . *must* capture the essence of political morality" to a new faith in "a doctrine of basic human rights, protecting specific basic liberties and interests of individuals."[41]

There may be a new faith, but liberal individualism is not a new development in moral and political theory. At least since Thomas Hobbes, liberal individualists have employed the language of rights to buttress moral and political arguments, and the Anglo-American legal tradition has incorporated this language. The language of rights has served on occasion as a means to oppose the status quo, to assert claims that demand recognition and respect, and to promote social reforms that aim to secure legal protections for individuals. Historically this language was instrumental in securing certain freedoms from established orders of religion, society, and state, such as freedom of the press and freedom of religious expression.

The vital role of civil, political, and legal rights in protecting the individual from societal intrusions is now beyond serious dispute, but the idea that rights provide the fountainhead for ethical and political theory has been strongly resisted (for example, by many utilitarians and Marxists). Individual interests are often at odds with communal or institutional interests. In discussions of health care delivery, for example, proponents of a broad extension of medical services often appeal to the "right to health care," whereas opponents sometimes appeal to the "rights of the medical profession." Many participants in these moral, political, and legal debates seem to presuppose that arguments cannot be made persuasive unless they can be stated in the language of rights, although other participants prefer to avoid the confrontational connotation of rights language.

The Nature and Status of Rights

Rights are justified claims that individuals and groups can make upon others or upon society. To have a right is to be in a position to determine, by one's choices, what others are to do or need not do.[42] Rights give us a claim based on a system of rules that authorize us to affirm, demand, or insist upon what is due. If a person possesses a right, others are validly constrained from interfering with the exercise of that right. Claiming will hereafter be understood as a rule-governed activity. The rules may be legal rules, moral rules, institutional rules, or rules of games, but all rights exist or fail to exist because the relevant rules either allow or disallow the claim or entitlement in question. These rules distinguish valid claims from invalid claims. *Legal* rights are claims that are justified by legal principles and rules, and *moral* rights are

claims that are justified by moral principles and rules. A right, then, is a justified claim or entitlement, validated by moral principles and rules.[43]

A rights holder need not assert his or her rights in order to have them. For example, small children, the comatose, and the mentally handicapped may not be able to claim their rights. Nonetheless, claims can be made for them by authorized representatives.

Absolute and Prima Facie Rights

Some rights may be absolute, such as the right to choose one's religion or to reject all religion, but typically rights are not absolute. Like principles of obligation, rights assert only prima facie claims (in the sense of "prima facie" introduced in Chapter 1). Some writers have asserted that rights are absolute, at least in restricted contexts. Ronald Dworkin is well known for his view that rights are individuals' "political trumps" and cannot be overridden to advance social interests. Although political decisions normally advance communal interests, he argues that the whole point of rights language is to constrain the community from acting at the expense of individuals. However, as Dworkin recognizes, if the claims of public utility are highly significant, it is not justifiable to allow the individual to play a trump card.[44] Dworkin, then, advances a sound theory about the *purpose* of having rights, rather than about their stringency or absoluteness.

Legitimate conflicting rights must be balanced or specified to reduce the conflict. Even the right to life is not absolute, irrespective of competing claims or social conditions, as evidenced by common moral judgments about killing in war and killing in self-defense. We have a right not to have our lives taken without justification, not an absolute right to life. Any right can be legitimately exercised and can create obligations on others only if the right overrides competing rights. Rights such as a right to give an informed consent or refusal, a right to die, and a right to lifesaving medical technology must compete with other rights, often producing a need to further specify the rights or to balance competing claims.[45]

In light of this need for balance, a *violation* of a right should be distinguished from an *infringement* of a right.[46] Violation refers to an unjustified action against a right, whereas infringement refers to a justified action overriding a right. When a right is justifiably overridden, it is infringed but not violated.

Positive Rights and Negative Rights

Whereas a positive right is a right to be provided with a particular good or service by others, a negative right is a right to be free from some action taken by others. A person's positive right entails another's obligation to do something for that person; a negative right entails another's obligation to refrain from doing something.[47] Examples of both sorts of rights are found in biomedical practice, research, and policy. If a right to health care exists, for example, it is a positive right to goods and services grounded in a claim of justice (see Chapter 6, pp. this 348–358). However, the right to forgo a recommended surgical procedure is a negative right grounded in the principle of respect for autonomy. The liberal individualist tradition has generally found it easier to justify negative rights, but the recognition of welfare rights in modern societies has extended the scope of rights to positive rights.

Confusion about public policies governing biomedicine can often be traced to a failure to distinguish positive and negative rights. One example involves the U.S. Supreme Court decisions on abortion. Those who contend that the various abortion decisions are inconsistent fail to see that the Court first recognized a negative right and later refused to recognize a positive right. The Court first ruled that a woman's right to privacy gives her a right to have an abortion prior to fetal viability (and after fetal viability if her life or health is threatened). The constitutionally protected right of privacy is here construed exclusively as a negative right that limits state interference. Many people thought that the Court had concomitantly recognized a positive right in its early decisions, namely a right to receive aid and assistance. They were surprised when the Court later ruled that the federal and state governments do not have obligations to provide funds for nontherapeutic abortions.[48] The Court's reasoning is consistent. It affirms a negative right and denies a positive right. (Our analysis is limited to this issue of consistency. We are not evaluating the substance of the court decisions.)

This controversy, and rights generally, should be analyzed by reference to the distinction between the statements (1) "X has a right to do Y" and (2) "X acts rightly in doing Y." The distinction is between rights (or a right) and right conduct, as well as between rights and their right exercise.[49] Sometimes when we say that a person "has a right to do X," we mean that he or she does not do wrong in performing X. But often our statement that someone "has a right to do X" implies nothing about the morality of the act, other than that others have no right to interfere with it. Thus, one can consistently affirm that a woman has a moral or a legal right to have an abortion and likewise affirm that she is not acting rightly in exercising her right.

The Correlativity of Rights and Obligations

How are rights connected to obligations? To answer this question, consider the meaning of "X has a right to do or have Y." X's right entails that some party has an obligation either not to interfere if X does Y or to provide X with Y. If a state has an obligation to provide goods such as food or health care to needy

citizens, then any citizen who meets the relevant criteria of need can claim an entitlement to food or health care. This analysis suggests a firm but untidy *correlativity* between obligations and rights.[50]

Suppose a physician agrees to take John Doe as a patient and commences treatment. The physician incurs an obligation to Doe, and Doe gains correlative rights. There may be rights to a certain level of care and rights in care, such as the right to refuse treatment. This correlativity of rights and obligations is untidy because one use of the words *requirement*, *obligation*, and *duty* suggests that obligations do not always imply corresponding rights. For example, although we sometimes refer to requirements or obligations of charity, no person can claim another person's charity as a matter of right. If such norms express what we "ought to do," they do so not from obligation but from personal ideals that exceed obligation. These commitments are best construed as self-imposed "oughts" that are not required by morality and that do not generate rights-claims for other persons.

A traditional distinction between obligations of perfect obligation and obligations of imperfect obligation can help us analyze this problem. Justice exemplifies perfect obligation, which entails a correlative right; whereas kindness, generosity, and charity exemplify imperfect obligation, which entails no correlative right. Mill argued that "Justice implies something which is not only right to do, and wrong not to do, but which some individual person can claim from us as his moral right. No one has a moral right to our generosity or beneficence, because we are not morally bound to practice those virtues towards any given individual."[51] Mill rightly saw that obligations of justice have correlative rights and are perfect. But, as we explicate beneficence in Chapter 5, many obligations of beneficence are also perfect obligations. We therefore need to augment Mill's analysis (here using beneficence as an example): (1) Some obligations of beneficence are *perfect* (for example, the obligations of rescue discussed in Chapter 5 and parental obligations to protect children), and (2) some obligations of beneficence are *imperfect* (for example, kindness and generosity), just as Mill describes them. But (3) some so-called "obligations" of beneficence are *self-imposed* requirements that are neither perfect nor imperfect obligations (for example some forms of kindness and generosity). For type 1, perfect obligations, the correlativity thesis always holds; these obligations and rights are those we typically find proper for enforcement by moral and legal sanctions, because a violation of rights and a failure of obligation are involved. Self-imposed requirements of type 3, by contrast, are optional and never have correlative rights. Obligations of type 2 may or may not have correlative rights. (We examine these problems further in Chapters 5 and 8.)

It is sometimes unclear without additional specification which obligation is correlative to a right, although it is clear that *some* obligation is correlative. Consider again the right to life. "X has a right to life" means that the moral system (or the legal system) imposes an obligation on other persons not to deprive X of life. However, this right does not specifically entail that X cannot come to an agreement with another party to end X's life through an act of euthanasia. What X *wants* makes a difference to how we understand rights, waivers of rights, and the exercise of rights. We conclude that rights language is correlative to obligation language, but in an untidy way requiring careful attention to particular contexts and often further specification of both rights and their correlative obligations.

The Primacy of Rights

The correlativity thesis does not determine whether rights or obligations, if either, is the more fundamental or primary category. The proposal that ethical theory should be "rights-based"[52] springs from a conception of the function and justification of morality. If the function of morality is to protect individuals' interests (rather than communal interests), and if rights (rather than obligations) are our primary instruments to this end, then moral action-guides are rights-based. Rights thus precede obligations and any other forms of protection.

This proposal can be illustrated by a theory we encounter in Chapter 6: the libertarian theory of justice. One representative, Robert Nozick, maintains that "Individuals have rights, and there are things no person or group may do to them (without violating their rights)."[53] He takes the following rule to be basic in the moral life: All persons have a right to be left free to do as they choose. The obligation not to interfere with this right follows from the right itself. That it "follows" is an indication of the priority of a rule of right over a rule of obligation. That is, an obligation is derived from a right.

Another rights-based argument that uses *positive or benefit* rights has been advanced by Alan Gewirth:

> Rights are to obligations as benefits are to burdens. For rights are justified claims to certain benefits, the support of certain interests of the subject or right-holder. Obligations, on the other hand, are justified burdens on the part of the respondent or duty-bearer; they restrict his freedom by requiring that he conduct himself in ways that directly benefit not himself but rather the right-holder. But burdens are for the sake of benefits, and not vice versa. Hence obligations, which are burdens, are for the sake of rights, whose objects are benefits.
>
> Rights, then, are prior to obligations in the order of justifying purpose . . . in that respondents have correlative obligations *because* subjects have certain rights.[54]

These rights-based accounts do not reject the correlativity thesis. Rather, they accept a priority thesis holding that obligations follow from rights, not the converse. Rights form the justificatory basis of obligations because they best capture the purpose of morality, which is to secure liberties or other benefits for a rights-holder.

A Critical Evaluation of Liberal Individualism

Problems with rights-based theories. One problem with basing ethics in rights is that rights are only a piece of a more general account that stakes out what makes a claim valid. Justification of the system of rules within which valid claiming occurs is not itself rights-based. Pure rights-based accounts also run the risk of truncating or impoverishing our understanding of morality, because rights cannot account for the moral significance of motives, supererogatory actions, virtues, and the like. Such a limited theory would fare poorly under criteria of comprehensiveness and explanatory and justificatory power. Accordingly, rights-based accounts should not be understood as a comprehensive or complete moral theory, but rather as an account of the minimal and enforceable rules that communities and individuals must observe in their treatment of all persons.

Normative questions about the exercise of rights. Often the question is not whether someone has a right, but whether the right possessed should or should not be exercised. If a person says, "I know you have the right to do x, but you should not do it," this moral claim cannot be reduced to a statement of a right. One's obligation or character, not one's right, is in question. Even if we had a full and complete theory of rights, we would still need a theory of obligation, at least about the appropriate exercise of rights, and it does not appear possible to develop a satisfactory account by attention only to rights and their limits.

The neglect of communal goods. Liberal individualists sometimes write as if the major concern of social morality is the protection of individual interests against government intrusion. This vision is too limited, because it excludes not only bona fide communal demands and group interests, but also communal goods and forms of life such as public health, biomedical research, and the protection of animals. The better perspective is that social ideals and principles of obligation are as critical to social morality as rights, and that neither is dispensable. Rights can sometimes be overridden by momentous communal interests.

The adversarial character of rights. Finally, the language of *claims* and *entitlements* is often unnecessarily adversarial. For example, the current interest in children's rights gives children many vital protections against abuse (for instance, when parents refuse to authorize lifesaving therapies for children for inappropriate reasons), but the notion that children have claims against their parents is an inadequate framework to express the moral character of the parent–child relationship. The attempt to understand this relationship and others

such as health care relationships strictly in terms of rights neglects and may even undermine the affection, sympathy, and trust at the core of the relationship. This is not to suggest that rights are inherently adversarial or that they are dispensable, but rather to note that rights theory is a partial framework.

A Constructive Evaluation of Liberal Individualism

In recent ethical theory some writers have sought to replace the language of rights altogether. The thought is either that rights language can be replaced by another vocabulary (obligations, virtues, etc.) or that the assertion of valid individual claims against society has risky implications. We reject such views, and we accept both the correlativity thesis and the moral and social purposes served by traditional interpretations of basic human rights.

We suspect that no part of the moral vocabulary has done more to protect the legitimate interests of citizens in political states than the language of rights. Predictably, injustice and inhumane treatment occur most frequently in states that fail to recognize human rights in their political rhetoric and documents. As much as any part of moral discourse, rights language crosses international boundaries and enters into treaties, international law, and statements by international agencies and associations. Rights thereby become acknowledged as international standards for the treatment of persons and the evaluation of communal action.

Being a rights-bearer in a society that enforces rights is both a source of personal protection and a source of dignity and self-respect. By contrast, to maintain that someone has an *obligation* to protect another's interest may leave the recipient in a passive position, dependent upon the other's good will in fulfilling the obligation. When persons possess enforceable rights correlative to obligations, they are enabled to be active, independent agents pursuing their projects and making claims. What we often cherish most is not that someone is obligated to us, but that we have a right that secures for us the opportunity to pursue and claim as ours the benefit or liberty that we value.

Communitarianism: Community-Based Theory

Communitarian theories view everything fundamental in ethics as deriving from communal values, the common good, social goals, traditional practices, and the cooperative virtues. Conventions, traditions, and social solidarity play a far more prominent role in communitarian theories than in the types of theory discussed to this point.

How might communitarians approach the case of potential kidney transplantation discussed previously in this chapter? Their first inquiry would not be which rights are at stake, but which communal values and relationships are

present or absent. They would focus on the family as a small community intermediate between the individual and the state. They would likely ask which acts, rules, and policies of living organ donation, privacy, and confidentiality best reinforce and promote communal values, including family values.

Communitarian critics of the father's behavior, which reduces his daughter's chances of survival, would charge that he is insufficiently committed to the goods of the family and presupposes the values of liberal individualism in standing on his rights, without adequately attending to his responsibilities. Critics would likely see the father as a twisted product of a society that focuses too much on protecting rights such as autonomy and privacy, and they may view the physician in the same light. The physician would certainly be expected to consider whether his actions conform to traditions of medicine, with its communal goods, codes, and virtues. In this tradition, deception has often been justified in the treatment of the patient, but the father requests that others be deceived, a relatively rare request and one with less clear historical precedent in medical practice. By contrast, nondisclosure to others because of confidentiality does have clear historical precedent in medicine, but these rules are not absolute and have often been overridden by a larger social interest.

The communitarian will support actions that express communal values as well as actions having a positive impact on a community. The father contends that if the physician tells other members of the family the true reasons for his decision not to donate, it would wreck the family. The father's prediction about this negative impact may or may not be correct, but what his actions express about his own lack of commitment to the family's welfare is notable. From the communitarian's perspective, the father embodies the vices of liberal individualism rather than the cooperative virtues.

The Repudiation of Liberalism

Contemporary communitarians repudiate central tenets in what is often called *liberalism*, a term that is defined through cardinal premises in the types of theory we have discussed in three previous sections: utilitarian, Kantian, and liberal individualist theories. What makes them jointly "liberal" is their commitment to what Mill defended as *individuality*, what Kant called *autonomy*, and what liberal individualists protect as *rights of the person*. Each type of theory protects the individual against the state, and—on the communitarian interpretation—each also asserts that the state should neither reward nor penalize different conceptions of the good life held by individuals. Postulates of individual autonomy, rights against the state, and community neutrality toward conflicting values, then, are the central elements of liberalism to which communitarians object.

In reacting critically to liberalism, contemporary communitarians repudiate

the *theory* as well as *current societies* established on the premises of liberal theory, including many contemporary Western political states.[55] According to communitarians, these societies lack a commitment to the general welfare, to common purposes, and to education in citizenship, while expecting and even encouraging social and geographic mobility, distanced personal relations, welfare dependence, breakdowns in family life and marital fidelity, political fragmentation, and the like. The number of abandoned children and elderly parents, social and familial fragmentation, the disappearance of meaningful democracy, and the lack of effective communal programs are, according to communitarians, the disastrous products of liberalism.

The meaning of *community* and its synonyms varies. Some communitarians refer almost exclusively to the political state as the community, whereas others refer to smaller communities and institutions with defined goals and role obligations. Some include the family as a basic communal unit, within which being a parent and being a child involve specific roles and responsibilities. Much of what one ought to do in communitarian theories is determined by the social roles assigned to or acquired by a person as a member of the community. Understanding a particular system of moral rules, then, requires an understanding of the community's history, sense of cooperative life, and conception of social welfare.

With regard to theory, communitarian criticisms have often been directed at Mill and Kant, but recently they have been aimed at Rawls, whose liberal principle that the rights of individuals cannot legitimately be sacrificed for the good of the community has been a particular target of communitarian censure.[56] These communitarian criticisms of liberal theories seem to amount to the following: Liberalism (1) fails to appreciate the constructive role of the cooperative virtues and the political state in promoting values and creating the conditions of the good life, (2) fails to acknowledge shared goals and obligations that come not from freely made contracts among individuals, but from communal ideals and responsibilities, and (3) fails to understand the human person as historically constituted by and embedded in communal life and social roles.

Michael Sandel describes the positive aspect of communal life that is allegedly missed by liberal theory:

> In so far as our constitutive self-understandings comprehend a wider subject than the individual alone, whether a family or tribe or city or class or nation or people, to this extent they define a community in the constitutive sense. And what marks such a community is not merely a spirit of benevolence, or the prevalence of communitarian values, or even certain "shared final ends" alone, but a common vocabulary of discourse and a background of implicit practices and understandings.[57]

Communitarians thus revitalize Hegel's criticism of Kant that was mentioned in Chapter 1 (namely, Kant presents an "empty formalism" without an "immanent doctrine of duties") and apply it to liberals: They miss the essence of

UNIVERSIDAD RAFAEL LANDIVAR BIBLIOTECA

morality by emphasizing abstract principles and abstract agents, while failing to see that both principles and agents are social products of communal life. Communitarians also propose that we give up the principles, politics, and language of rights in favor of the principles, politics, and language of the common good and the community's way of life.[58]

Militant and Moderate Forms of Communitarianism

Communitarianism can be distinguished into *militant* and *moderate* forms. Militants firmly support community control and reject liberal theories. This approach has been supported by influential contemporary moral, social, and political thinkers, including Alasdair MacIntyre, Charles Taylor, and Michael Sandel. By contrast, moderates emphasize the importance of various forms of community—including the family and the political state—while attempting to accommodate rather than reject strands in liberal theories. This sense of communitarianism includes figures as diverse as Aristotle, Hugo Grotius, David Hume, G. W. F. Hegel, John Mackie, and Michael Walzer. For them social order and morality rest on historically developed norms, and moral rules derive their acceptability and correctness from these shared conventions. Although *communitarianism* is a recently coined term typically used for the militant form, we will use it for both forms. We will criticize militant theories, while relying on the moderate theories for our constructive evaluation.

Militant communitarianism is hostile to rights, sees liberalism as "born of antagonism to all tradition," and aims to perpetuate and even impose on individuals conceptions of virtue and the good life that limit the rights conferred by liberal societies. These communitarians see persons as intrinsically *constituted* by communal values and as best suited to achieve personal goods through communal life.[59] In addition, MacIntyre argues that we have inherited many incoherent fragments of once coherent schemes of thought and action, and only if we understand our peculiar historical and cultural situation can we recognize the problematic dimensions of the enterprise of moral evaluation and moral theory.[60]

The moderate communitarian takes a stance far less opposed to autonomy and individual rights. A typical example is J. L. Mackie's appeal to "intersubjective standards," meaning that community-wide agreements form the basis of acceptable moral rules and that these intersubjective agreements cannot be further validated or invalidated by appeal to rationality. Mackie understands morality entirely in terms of social practices that express what is demanded, allowed, enforced, and condemned in the community. Nonetheless, he insists that moral judgments need not be viewed as unchanging conventional rules beyond the possibility of reform: "Of course there have been and are moral heretics and moral reformers. . . . But this can usually be understood as the extension, in ways which, though new and unconventional, seemed to them to be required for consistency of rules to which they already adhered as arising out of an existing way of life."[61]

The Primacy of Social Practices

Alasdair MacIntyre and other communitarians have traced to Aristotle the thesis that local community practices and their corresponding virtues should have priority over ethical theory in normative decisionmaking. MacIntyre uses "practice" to designate a cooperative arrangement in pursuit of goods that are internal to a structured communal life. Social roles of parenting, teaching, governing, healing, and the like involve practices. "Goods internal to a practice" are achievable, according to MacIntyre, only by engaging in the practice and conforming to its constraints and standards of excellence. In the practice of medicine, for example, goods internal to the profession exist, and these determine what it is to be a good physician. The virtues of physicians flow from communal and institutional practices of care, practical wisdom, and teaching. Medicine, like other professions and political institutions, has a history that sustains a tradition requiring participants in the practice to cultivate certain virtues.[62]

The importance of traditional practices and the need for communal intervention to correct socially disruptive outcomes are standard themes in communitarian thought. For example, Sandel proposes that we disallow plant closings that devastate local communities and that we ban pornography when it deeply offends a community's way of life.[63] As an example of communitarians' promotion of the common good in biomedical ethics, consider their debate with liberal individualists over policies of obtaining cadaveric organs for transplantation. Based on principles of liberal individualism, but with an interest in obtaining cadaveric organs to save lives, all states in the United States adopted the Uniform Anatomical Gift Act in the late 1960s and early 1970s. This act gives individuals the right to make decisions about the donation of their organs through a donor card. If the individual has not made a decision prior to death, the law authorizes the family to decide whether to donate the decedent's organs. On the basis of opinion polls, it was expected that individuals would sign donor cards and provide a sufficient supply of organs, thereby avoiding the need to search for living donors of kidneys.

In practice, however, few individuals sign donor cards, the cards are rarely available at the time of death, and procurement teams virtually always check with the family even if the decedent left a valid donor card. As a result, a communitarian focus has emerged. The family has become the primary donor (that is, the decisionmaker about donation) rather than the individual, and because the supply of organs has remained limited, various policies have been

considered and some adopted that aim to promote the common good more vigorously. Even approaches that protect individual rights attempt to educate people about the need for organs, and some propose *requiring* people to make a decision about donation, for instance, when obtaining a driver's license. Laws and regulations have also been implemented to require hospitals to ask families whether they know the wishes of the decedent and want to donate the decedent's organs.

Some communitarians now recommend still stronger laws to make organ procurement a well-defined community project rather than a matter of individual or even family decisions. They defend *presumed consent* laws, which would parallel laws in several states for corneas and laws in several countries for solid organs. These laws presume that individuals or families have decided to donate unless they have registered a dissent. A more stringent proposal is the routine salvaging of organs unless objections are registered. Here communitarians defend a policy of organ retrieval on grounds that members of a community should be willing to provide others objects of lifesaving value when no cost to themselves is required.[64] A few commentators even recommend harsher policies of conscription of cadaveric organs to reflect community ownership of cadaveric body parts. The latter approach conflicts so deeply with liberal individualistic values that it has not received serious consideration. Nevertheless, an extreme alternative approach based on liberal individualism, a market in organs, has been declared illegal in the United States because of concern about exploitation and coercion. Proponents of a market in organs typically view their communitarian opponents as zealous and inconsistent, because they allow and encourage individual or family *gifts* to benefit others but rule out *sales* that would provide the same benefit and perhaps even increase the supply of organs for transplantation.[65]

An emphasis on the community and the common good also appears in debates about the allocation of health care. In Daniel Callahan's communitarian account, we should enact public policy from a shared consensus about the good society, not on the basis of individual rights. Liberal assumptions about state neutrality should be scrapped, and society should be free to implement a substantive concept of the good. According to Callahan, biomedical ethics should use communitarian values to implement or revise social laws and regulations governing the promotion of health, the use of genetic knowledge, the use of advances in medical technology, responsibilities to future generations, and the limits of health care for the elderly. In each case, the question to be asked is, "What is most conducive to a good society?," not "Is it harmful or does it violate autonomy?"[66] Here we see a close similarity to utilitarian proposals. However, communitarians typically reject the principle of utility on grounds that it is remote from actual communal decisionmaking and, in any event, is

individualistic in its effort to sum up individual benefits and costs for public policy.

Although many communitarians critique and propose specific acts, practices, and policies, such as procuring organs or allocating health care, few systematic communitarian proposals have emerged for biomedical ethics as a whole. One exception is Ezekiel Emanuel's vision of medical ethics, which rests on the following claims: The ends of medicine, as affirmed by the profession, have been shaped by public laws and public values. These ends are understood through a framework of shared political convictions, conceptions of justice, and ideas of the good life. In place of the liberalism that has typically undergirded medical ethics, Emanuel proposes a moderate communitarianism closely connected to political theory. This communitarianism is moderate by virtue of its acceptance of pluralistic conceptions of the good life and its recognition of some individual rights. Yet it remains communitarian because democratic initiatives will be needed to fashion a community's conceptions of the good life into policies and laws. Emanuel envisions thousands of community health plans in which citizen-members deliberate about conceptions of the good life and debate policies such as those for termination of life-sustaining treatment for incompetent patients and the allocation of medical resources.[67]

A Critical Evaluation of Communitarian Ethics

Several claims by militant communitarians rely on questionable accusations and arguments. We will concentrate on these problems in our criticisms. However, many themes in moderate communitarianism are unproblematic and even acceptable to many advocates of liberal theories. We focus on these unproblematic positions in our constructive section below.

An unfair account of liberal theories. Militant communitarians suggest that liberal theorists defend atomic, isolated individuals and have a corrupting skepticism about communal goods.[68] This characterization is inaccurate and unfair. Mill and Rawls, the figures most frequently attacked by communitarians, never depict either individuals or the communal good in these terms, and both philosophers develop a theory of the common good, as well as an account of social traditions and political community.[69] Mill thought he had captured how historical traditions converge to the principle of utility, which he construed as a principle of communal welfare. Even in *On Liberty*, Mill argued that a community should take steps to ensure adequate public discussion of what constitutes the good of the community. Liberty functions in his arguments to protect individuals against mistakes in planning communal pursuits of the good, and he defends individuality *because* it conduces to a constantly readjusted and improved so-

cial unit. Rawls defends rights and the value of liberty in society, in part, because social ends can be corrected better in an open society than in a society controlled by tradition.[70]

A false dichotomy: community or autonomy. Communitarians present us with two false dichotomies: (1) either liberal accounts of rights and justice have priority or the communal good has priority,[71] and (2) either radical autonomy in decisionmaking is protected or communal determination of social goals is protected against the individual. A more accurate picture is that we inherit various social roles and goals from traditions. We then critique, adjust, and attempt to improve our beliefs over time through free discussion and collective arrangements. Individuals and groups alike progressively interpret, revise, and sometimes even replace traditions with new conceptions that adjust and foster community values. This outlook of liberalism is, as Joel Feinberg notes, entirely compatible with communal interests: "It is impossible to think of human beings except as part of ongoing communities, defined by reciprocal bonds of obligation, common traditions, and institutions. . . . The ideal [in liberals' accounts] of the autonomous person is that of an authentic individual whose self-determination is as complete as is consistent with the requirement that he is, of course, a member of a community."[72]

A failed challenge to rights. Communitarians sometimes argue against rights (especially natural rights) on grounds that they do not exist.[73] At other times they argue against rights on grounds that rights stall communal organization and dull our sense of social union. Both claims miss the valuable consequences that rights have for communities. We value rights because, when enforced, they provide protections against unscrupulous behavior, promote orderly change and cohesiveness in communities, and allow diverse communities to coexist peacefully within a single political state.[74] As Judith Jarvis Thomson notes,

> How much more satisfying the life in an 'organic community' than the life of alienation in a modern state! The ideal of the hive is seductive and fuels all communitarian ideologies. But the bee-like creatures of our hive-like world are not in fact kind to each other; each is indifferent to the others except insofar as the others are parts of the whole. . . . The ideal of the state as hive cannot be made real: it is amazing that communitarians have expected otherwise.[75]

Even if we grant communitarian arguments that the best life is communal life, it would not follow that communities should determine the individual's goals or truncate individual rights. The major reason for the prominence of rights in moral and political theory is that they stand as a shield against communal intrusion by governments. This and similar criticisms raise profound questions about how well communitarianism fares on several of the criteria for theory construction presented in the beginning of this chapter, especially output, explanatory, and justificatory power.

A Constructive Evaluation of Communitarianism

By emphasizing historical traditions and institutional practices, communitarian theories have made a substantial contribution to the redirection of ethical theory in recent years, and have also helped us rediscover the importance of community even if we accept liberal values. Communitarians rightly emphasize the need to foster neighborhood associations, create communal ties, promote public health, and develop national goals. Also to be welcomed is the return in some communitarian theories to such landmarks in ethical theory as the writings of Aristotle, Hume, and Hegel. These more community-minded philosophers deserve status as great classical theorists, alongside Mill and Kant.

Ethics of Care: Relationship-Based Accounts

Another family of moral reflections is widely referred to as the *ethics of care.* It shares some premises with communitarian ethics, including some objections to central features of liberalism and an emphasis on traits valued in intimate personal relationships, such as sympathy, compassion, fidelity, discernment, and love. *Caring* in these accounts refers to care for, emotional commitment to, and willingness to act on behalf of persons with whom one has a significant relationship. Noticeably downplayed are Kantian universal rules, impartial utilitarian calculations, and individual rights.

Proponents of an ethics of care would approach the case we have been examining by focusing on relationships involving care, responsibility, trust, fidelity, and sensitivity. The father who elects not to donate a kidney expresses some concern about his daughter's suffering, but his response is arguably grounded mainly in concern about himself. He does not think he can justify his behavior to his wife, who will, he believes, distrust him and "accuse him of allowing his daughter to die." Even if we give the father the benefit of the doubt about motives and trustworthiness, whether his care is responsibly expressed in donation or nondonation will depend in part on the balance of risks and benefits and his courage in confronting the risks.

The physician in this case faced several conflicts within relationships of care—to the dying daughter, her siblings, the reluctant father, the mother, and the family as a unit. Just as many moral theories face conflicts of principles and rights, the ethic of care faces conflicts among responsibilities in such situations. Traditional moral theory has typically concentrated on answers to questions about whether to lie or break confidentiality. The ethic of care, by contrast, emphasizes that it is not only important what the physician does—for example,

breaks or maintains confidentiality—but also how actions are performed, which motives underlie them and whether positive relationships are promoted or thwarted. The trustworthiness of the physician and the quality of his care and sensitivity in the face of the father's unusual request for deception are all integral moral elements from the perspective of the ethics of care.

Two Speakers in a Different Voice

The origin of the ethic of care was predominantly in feminist writings. The themes included how women display an ethic of care, by contrast to men, who predominantly exhibit an ethic of rights and obligations. We begin with two figures who have played prominent roles in this recent history, psychologist Carol Gilligan and philosopher Annette Baier.

Gilligan's psychological account. The hypothesis that "women speak in a different voice"—a voice that traditional ethical theory has drowned out—arose in Gilligan's book, *In a Different Voice*. She maintained that women's moral development is typically distinct from men's, a fact she thought disregarded by influential psychological studies of moral development whose conceptions were based on studies of males only. She claimed to discover "the voice of care" through empirical research involving interviews with girls and women. This voice, she said, stresses empathic association with others, not based on "the primacy and universality of individual rights, but rather on . . . a very strong sense of being responsible." In her studies, female subjects typically view morality in terms of responsibilities of care deriving from attachments to others, whereas male subjects typically see morality in terms of rights and justice. Men look to and are formed by freely accepted relationships and agreements; women look to and are formed by contextually given relationships such as those of the family.[76]

Gilligan, then, identified two modes of relationship and two modes of moral thinking: an ethic of care in contrast to an ethic of rights and justice. She does not claim that these two modes of thinking are strictly correlated with gender or that all women or all men speak in the same moral voice.[77] Rather, she believes that men *tend* to embrace an ethic of rights using quasi-legal terminology and impartial principles, accompanied by dispassionate balancing and conflict resolution, whereas women *tend* to affirm an ethic of care that centers on responsiveness in an interconnected network of needs, care, and prevention of harm. Taking care of others is the core notion, and it is modelled on relationships such as those between parent and child.[78]

Baier's philosophical account. Gilligan's interpretation of empirical data has parallels in philosophical ethics. In Annette Baier's account, the reasoning and

methods of women who write in ethical theory is noticeably different from traditional theories. She claims to hear in contemporary female philosophers, despite their diversity, the same different voice that Gilligan heard in her studies, but one made "reflective and philosophical."[79] She deplores the near-exclusive emphasis in modern moral philosophy on universal rules and principles, and she sternly rejects Kantian contractarian models with their emphasis on justice, rights, law, and particularly autonomous choice among free and equal agents. The conditions of social cooperation, especially in families and in communal decisionmaking, are, Baier observes, typically unchosen and intimate, and they involve unequals in a relational network. Her thesis is not that traditional ethical theories are false or even outmoded, but that they capture only a piece of the larger moral world.[80]

Baier envisions not a grand system of ethics that holds together all the diverse strands, but smaller scale systems that pull together a few strands. In casting about for a connecting bridge to span an ethic of love with an ethic of obligation, she proposes "appropriate trust" as a bridging concept. She does not recommend that we discard categories of obligation, but that we make room for an ethic of love and trust, including an account of human bonding and friendship. Traditional models of ethical theory often fail to acknowledge how parents and health care professionals, for example, see responsibilities to their children and patients in terms of care, loving attachment, meeting needs, and providing sustenance.[81]

Criticisms of Traditional Liberal Theories

Proponents of the care perspective offer a direct challenge to liberal values. Two criticisms of liberalism deserve special mention.[82]

Challenging impartiality. According to the care perspective, liberalism has lost sight of the full sweep of morality by taking a standpoint of detached fairness. This orientation is suitable for some moral relationships, especially those in which persons interact as equals in a public context of impersonal justice and institutional constraints. But lost in this *detachment* is an *attachment* to that which we care about most and which is closest to us—for example, our loyalty to groups. In the absence of public and institutional constraints, partiality toward others is not only morally permissible but is the expected norm of interaction and is an ineliminable feature of the human condition. Without exhibiting partiality we stand to sever important relationships and to alienate others. In seeking a blinded impartiality, liberalism risks making us blind and indifferent to the special needs of and relationships with others. Although impartiality is a moral virtue in some contexts, it is a moral vice in others. This two-sidedness is overlooked in traditional liberal theory, which simply aligns good and mature

moral judgment with moral distance.[83] The care perspective is especially meaningful for roles such as parent, friend, physician, and nurse, in which contextual response, attentiveness to subtle clues, and the deepening of special relationships are likely to be more momentous morally than impartial treatment.

Challenging universal principles. An aversion to abstract principles, the instruments of impartiality, is also characteristic of the ethics of care. As long as principles allow room for discretionary and contextual judgment, the ethics of care need not dispense with principles. However, like many proponents of virtue theory, defenders of the ethics of care find principles often irrelevant, unproductive, ineffectual, or constrictive in the moral life. A defender of principles could say that *principles* of care, compassion, and kindness tutor our responses in caring, compassionate, and kind ways. But this claim seems hollow. Our moral experience suggests that our responses rely on our emotions, our capacity for sympathy, our sense of friendship, and our knowledge of how caring people behave.

Consider, as an example, the following report by physician Timothy Quill and nurse Penelope Townsend of a discussion with a young woman who has just been told that she is HIV infected:[84]

PATIENT: Oh God. Oh Lord have mercy. . . . Please don't do it again. Please don't tell me that. Oh my God. Oh my children. Oh Lord have mercy. Oh God, why did He do this to me? . . .

DR QUILL: First thing we have to do is learn as much as we can about it, because right now you are okay.

PATIENT: I don't even have a future. Everything I know is that you gonna die anytime. What is there to do? What if I'm a walking time bomb? People will be scared to even touch me or say anything to me.

DR QUILL: No, that's not so.

PATIENT: Yes they will, 'cause I feel that way . . .

DR QUILL: There is a future for you . . .

PATIENT: Okay, alright. I'm so scared. I don't want to die. I don't want to die, Dr Quill, not yet. I know I got to die, but I don't want to die.

DR QUILL: We've got to think about a couple of things. . . .

Quill and Townsend have moral responsibilities to their patient, but it is difficult to capture their responsibilities through principles and rules. We can produce rough generalizations about how caring physicians and nurses respond to patients, for example, but these generalizations will not be subtle enough to give helpful guidance for the next patient. Each situation calls for a set of

responses exceeding what the generalization captures, and behavior that in one context is caring seems to intrude on privacy or to be offensive in another setting.

Relationship and Emotion

Two constructive themes are central to the ethics of care: mutual interdependence and emotional response.

Mutual interdependence in relationships. The ethics of care maintains that many human relationships—for example, in health care and research—involve persons who are vulnerable, dependent, ill, and frail and that the desirable moral response is attached attentiveness to needs, not detached respect for rights. Feeling for and being immersed in the other person establish vital facets of the moral relationship. Accordingly, this approach features responsibilities that a rights-based account may ignore in the attempt to protect persons from invasion by others.[85]

A role for the emotions. Ethical theory since the late eighteenth century has exhibited a cognitivist proclivity; that is, it has regarded theory and moral judgment as the affairs of reason, rather than of emotion or passion. Kant joined many other writers in the history of ethics, such as Plato, in depicting the emotions, feelings, passions, and inclinations as distracting impediments to moral judgment. These philosophers call for a struggle against desire, impulse, and inclination, in order that a more rational course of action will ensue. Actions done from desire, impulse, or inclination may be good in these theories, but not *morally* good, because they are not done from an appropriate cognitive framework.

The ethics of care corrects this cognitivist bias by giving the emotions a moral role. Having a certain emotional attitude and expressing the appropriate emotion in acting are morally relevant factors, just as having the appropriate motive for an action is morally relevant. The person who acts from rule-governed obligations without appropriately aligned feelings such as worry when a friend suffers seems to have a moral deficiency. In addition to expressing their feelings in their responses, agents also need to attend to the feelings of persons toward whom they act in moral relationships. Insight into the needs of others and considerate alertness to their circumstances often come from the emotions more than reason.[86] In the history of human experimentation, for example, those who first recognized that some subjects of research were being brutalized, subjected to misery, or placed at unjustifiable risk were persons who were able to feel compassion, disgust, and outrage through insight into the situation of these research subjects. They exhibited emotional discernment of

and sensitivity to the feelings of subjects, where others lacked comparable responses.

This emphasis on the emotional dimension of the moral life does not reduce moral response to emotional response. Caring clearly has a cognitive dimension as well, because it involves an insight into and understanding of another's circumstance, needs, and feelings. As Hume pointed out, emotions motivate us and tell us much about a person's character, but it is the understanding that directs us in choosing a path of action.

A Critical Evaluation of the Care Ethic

The ethics of care emphasizes engaged, contextual, and even passionate moral thinking. As long as both passion and dispassion are acknowledged, few if any crippling criticisms can be brought against the ethics of care. Nonetheless, some problems need attention.

Underdeveloped theory. If one takes seriously the eight criteria for theory construction developed at the beginning of this chapter, then the ethics of care seems to fall short on criteria such as completeness, comprehensiveness, and explanatory and justificatory power. Of course, there may be a bias in the criteria: Because this list grows out of *traditional* accounts of theory that are often opposed by proponents of the ethics of care, it might be expected to reach a negative judgment on the ethics of care, which explicitly departs from traditional theory. But the heart of the problem is the lack of a developed and integrated body of reflections to supply the concepts and connections needed to satisfy these criteria. As Baier has pointed out, the ethics of care needs one or more central concepts and a set of bridging concepts to link it to the legitimate concerns of traditional theory. The ethic of care, then, is an underdeveloped theory, but not necessarily an incorrect one.

Should impartiality be rejected? In deemphasizing justice, impartiality, rights, and obligations, the ethics of care must confront situations in which bona fide requirements of impartiality conflict with acting partially from care. Acting partially clearly must sometimes yield to acting impartially. On at least some occasions we need an impartial judgment to arbitrate between conflicting moral judgments or feelings.[87] It is doubtful that many who endorse the ethics of care want their theory to be interpreted so narrowly as to exclude all impartial judgments and considerations of justice and the public good. But a problem remains about whether the theory can successfully incorporate these moral notions without losing much of its critical thrust and uniqueness. The ethic of care, as Gilligan and others have defended it, recognizes that two perspectives exist, but can they can be made *coherent?* Alternatively, one might argue (as Nel Noddings does) that the ethics of care is the fundamental form of morality and that it is internally coherent. The latter, we suggest, simply gives up too much in the moral life.[88]

Too contextual and hostile to principles. One proponent of the ethics of care argues that in a defensible ethical theory, action should be "sometimes principle-guided, rather than always principle-derived."[89] This statement is a move in the right direction of coherence. However, if principles are accommodated by an ethics of care, does this inclusion undercut the grounds for antipathy to principles? The question is again one of coherence. Can the theory coherently trade on a rejection of some principles (Kantian principles, say), and at the same time accept a vital role for other principles (prima facie principles, say)?

We think principles will reappear in a more comprehensive theory and will enhance rather than weaken the ethics of care. If we agree that certain forms of sympathy and emotion are appropriate bases of motivation, we should be prepared for situations in which our actions are too partial and in need of correction by impartial principles. We are likely to judge more favorably persons who are close to us in intimate relationships, and yet on some occasions those who are distant from us deserve to be judged more favorably.

Feminist reservations about an ethics of care. Although initiated by feminist writers, the ethics of care has been sharply criticized by some feminists who worry that it attends to women's experiences as givers of care in traditional roles of self-sacrifice, but often neglects feminist insights into problems of oppression and dominance. Susan Sherwin argues that feminists should "be cautious about the place of caring in their approach to ethics; it is necessary to be wary of the implications of gender traits within a sexist culture. Because gender differences are central to the structures that support dominance relations, it is likely that women's proficiency at caring is somehow related to women's subordinate status."[90] She sees a need to examine the social context of care as well as to establish limits to the ethics of care. Both enterprises involve appeals to justice.

Without a broader framework, the ethics of care is too confined to the *private* sphere of intimate relationships and may serve to reinforce an uncritical adherence to traditional social patterns of assigning caretaker roles to women. Among health professionals, the ethic of care has been most widely appropriated by nurses. Without further explication, there is a danger that the ethics of care will be primarily located in nursing and primary care specialties in medicine to which many women are attracted, without having a major impact on health care as a whole.[91]

A Constructive Evaluation of the Care Ethic

The care ethic provides a needed corrective to two centuries of system-building in ethical theory and to the tendency to neglect themes such as sympathy, the moral emotions, and women's experiences. A morality centered on care and concern can potentially serve health care ethics in a constructive and balanced fashion, because it is close to the processes of reason and feeling exhibited in clinical contexts. We have seen that sympathy, friendliness, compassion, and trust cannot easily be brought under rules of behavior or even under a principle such as beneficence. Physician and nursing ethics have recently been presented in codes that express obligations and rights, but the ethics of care can retrieve basic commitments of caring and caretaking and help free health professionals from a narrow conception of their role responsibilities. Caring involves an open-minded responsiveness to another's needs as the other sees those needs, and therefore runs counter to the assumption that an established medical good will best meet those needs.

Disclosures, discussions, and decisionmaking in health care typically become a family affair, with support from a health care team. The ethics of care fits this context of relationships, whereas rights theory, for example, seems poorly equipped for it. Finally, correcting an undue obsession with impartiality requirements in traditional theories promises to have positive consequences because many aspects of character, forms of sensitivity, and modes of practical judgment exceed appeals to impartial principles. We return to these moral qualities in Chapter 8.

Casuistry: Case-Based Reasoning

Recently ethical theory has seen a revival of an approach with impressive influence in medieval and early modern philosophy. *Casuistry*, as it is called, focuses on practical decisionmaking in particular cases.[92] Casuists are skeptical of rules, rights, and theories divorced from history, precedent, and circumstance. Appropriate moral judgments occur, casuists say, through an intimate understanding of particular situations and the historical record of similar cases.[93]

Consider first how the casuist might approach the case of the father's refusal to become a donor. The casuist would begin by identifying particular features in the case rather than appealing to universal principles, utilitarian calculations, or rights. The casuist would then attempt to identify the relevant precedents and prior experiences with other cases, attempting to determine how similar and different this case is from experiences with other cases. In assessing what the father should do, the casuist would determine whether we typically insist, in relevantly similar cases, that parents bear comparable inconvenience and risk

to offer their children some chance of survival. In determining what the physician should do, analogous cases would be considered in which breaches of confidentiality are justified or unjustified. The objective is to act in light of any strong social consensus found in precedent cases in medicine and law. Such cases would indicate, for example, that physicians have a right and sometimes an obligation to breach confidentiality in order to prevent harm to others. Examples of these cases include reporting gunshot wounds and venereal diseases and in some contexts warning intended victims of a patient's threatened violence.

The casuist might also ask whether the father's refusal to donate would cause a *harm* to his daughter or would only *fail to benefit* her and whether a threatened or actual breach of confidentiality might be justified in an effort to force him to donate. Similarly, the casuist would ask whether a lie ("the father is not histocompatible") or a milder form of deception ("for medical reason the father should not donate") could be justified to prevent wrecking the family. The casuist would attempt to answer these questions by appeal to maxims grounded in experience and tradition, as well as by reasoning from analogous cases.

The Recent Recovery of Casuistry

The recent rise of casuistry has surprised many, because in the last three hundred years casuistry had fallen into a disrepute rivaling astrology.[94] To illustrate its former low repute, when *An Encyclopedia of Religion* was published in 1945, then-prominent philosopher Edgar Sheffield Brightman wrote the entry on "casuistry," which read (in full):[95]

> 1) The application of ethical principles to specific cases. 2) Quibbling, rationalization, sophistry or an attempt to justify what does not merit justification; this meaning is often associated with methods used by Jesuits. See equivocation.

This definition is still typical of entries in reference works. But contemporary casuists would argue that Brightman and mainstream critics have matters upside-down. Casuists claim that their approach is not an application of principles to cases—quite the reverse, it moves up from cases to principles—and is a system of justification that tries to surmount the sophistry of "applying" principles.

A Repudiation of the Mainstream in Modern Ethics

As with communitarian theories and the ethics of care, the casuist is motivated in part by a dissatisfaction with the dominant ethical theories, including Kantianism, utilitarianism, and rights theory. In particular, casuists dispute the use

of the model of scientific theory for ethical theory, the accompanying account of moral judgments, and the insistence on firm, universal principles.

Repudiating the model of a philosophical moral science. Some nineteenth- and twentieth-century philosophers seem to presuppose a model of a tidy, unified theory containing general and universal principles—in effect, a philosophical moral science.[96] Casuists reject this model, sometimes under the influence of Aristotle's conceptions of science and ethics. Aristotle noted the idea of a "first principle" that is certain and inherently justified belongs to science conceived on an axiomatic model,[97] but he held that principles in ethics are deeply embedded in the concrete world of human social conduct. Philosophers must obtain first principles by abstracting them from the mass of human actions and social practices.[98]

Casuists agree. Although there conceivably could be a first foundational principle for ethics that has absolute priority (not one from which all other moral content would follow), they maintain that moral beliefs and reasoning in fact do not follow this pattern. Ethics is not a demonstrative science, but a set of practices and types of judgment rooted in experience, wisdom, and prudence.

Repudiating moral judgment based on principles. Casuists interpret many moral philosophers to hold that cases are devoid of material that informs moral judgment, and thereby are powerless to determine obligation, blameworthiness, or praiseworthiness. Cases *illustrate* principles, *exemplify* dilemmas, *motivate* people to right actions, and the like; but cases are otherwise irrelevant to moral judgment. By contrast, casuists maintain that some forms of moral reasoning and judgment make no appeal to principles, rules, rights, or virtues. These forms include appeals to narratives, paradigm cases, analogies, models, classification schemes, and even immediate intuition and discerning insight.[99]

Rules and principles need not be excluded from moral thinking, but the casuist insists that moral judgments can be and often are made when no appeal to principles is possible. For example, we make moral judgments when principles, rules, or rights conflict and no further recourse to a higher principle, rule, or right is available. When principles are interpreted inflexibly irrespective of the nuances of the case, some casuists see a "tyranny of principles."[100] As a result, attempts at the resolution of moral problems suffer from a gridlock of conflicting principles, and moral debate becomes intemperate and interminable. This impasse can often be avoided, Albert Jonsen and Stephen Toulmin argue, by focusing on points of shared agreement about cases rather than on principles. The following is their prime example, drawn from their personal experiences during four years of work with the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research:

> The one thing [*individual* commissioners] could not agree on was *why* they agreed. . . . Instead of securely established universal principles, . . . giving them intellectual grounding for particular judgments about specific kinds of cases, it was the other way around.
>
> The *locus of certitude* in the commissioners' discussions . . . lay in a shared perception of what was specifically at stake in particular kinds of human situations. . . . That could never have been derived from the supposed theoretical certainty of the principles to which individual commissioners appealed in their personal accounts.[101]

In this account, casuistical reasoning rather than universal principles forged agreement. The commission functioned successfully by appeal to paradigms, particular cases, and families of cases, despite the various principles and inchoate moral theories individual commissioners held. Consensus about policies was reached by agreement on cases, when agreement would have been impossible to achieve on principles or theory. Although commissioners cited moral principles to justify their collective conclusions, Jonsen and Toulmin argue that these principles were less certain and central for commissioners in their deliberations than their particular judgments about cases.[102]

We agree that evidence supporting this interpretation exists in the work of the commission, but equally weighty evidence supports a justificatory role for principles in its deliberations, as we will see below.

Case-Based Reasoning and Judgment

Casuists typically hold that moral belief and knowledge evolve incrementally through reflection on cases, without essential recourse to a top-down theory. To support this thesis, casuists sometimes ask us to consider an analogy to case law. When the decision of a majority of judges becomes authoritative in a case, their judgments are positioned to become authoritative for other courts hearing cases with *similar* facts. This is the doctrine of precedent. Casuists see moral authority similarly: Social ethics develops from a social consensus formed around cases. This consensus is then extended to new cases by analogy to the past cases around which the consensus was formed. The underlying consensus and the paradigm cases become enduring and authoritative sources of appeal. For example, in the current literature of biomedical ethics, cases such as the Quinlan case, the Tuskegee Syphilis experiments, and the Quill case are constantly invoked not only to illustrate claims, but as sources of authority for new judgments.

As a history of similar cases and similar judgments mounts, we become more confident in our judgments. A "locus of moral certitude" is found in the judgments, and the stable elements crystallize in the form of tentative principles. As confidence in these generalizations increases, they are accepted less

tentatively and moral knowledge develops. Just as case law (legal rules) develops incrementally from legal decisions in cases, so the moral law (moral rules) develops incrementally.[103]

Casuists find "the essence of the casuistic mode of thinking" in a gradual movement from clear and resolvable cases to more complex and difficult cases. There is an "ordering of cases under a principle by paradigm and analogy." The process is similar to that of a physician in clinical diagnosis and recommendations. Paradigms of accurate diagnosis and proper treatment function as sources of comparison when new problem cases arise. Recommendations are made by analogy to the paradigm. If the analogy is proper, a resolution of the problem and a recommendation will be achieved, but if no close analogy is available, uncertainty will remain.[104]

Consider the following example (ours, not that of any casuist known to us): If a particular act of suicide is paradigmatically wrong, then it will have certain relevant similarities with other wrong actions of suicide. Again, if a particular act of suicide is justifiable, then it will share relevant features with other acts of suicide that are morally acceptable. When confronted with a case of *assisted* suicide by a physician, these analogous, settled cases will constitute primary (but not exhaustive) resources for reasoning about the new moral problem of assisted suicide. No principle about suicide or killing need be invoked in this process if the paradigms are sufficiently powerful.

What Role for Theory?

Casuists disagree among themselves about the value and limitations of theory in practical ethics. While some casuists are sharply critical of theory, others encourage theory construction as well as generalization from cases. Baruch Brody, for example, insists that ethical theory is both possible and desirable. Case-based judgment that rests on plausible intuition "is only the *first* stage in the process of coming to have moral knowledge. The *next* stage is that of theory formation. . . . The goal is to find a theory that systematizes these intuitions, explains them, and provides help in dealing with cases about which we have no intuitions. In the course of this systematization, it may be necessary to reject some of the initial intuitions on the grounds that they cannot be systematized into the theory."[105] This theory-accessible casuistry is more appealing than a casuistry that denounces or evades theory.[106]

A Critical Evaluation of Casuistry

Casuists have sometimes overstated the promise and output power of their account, while understating the value of competing accounts. These problems need to be corrected.

Problems of case interpretation and conflicting judgments. Casuists often write as if cases speak for themselves or inform moral judgment by their facts alone. Clearly they do not. *Interpretation* of cases is essential for moral judgment, and principles and theory typically play a legitimate role in the interpretation. For the casuist to move constructively from case to case, some recognized rule of moral relevance must connect the cases. The rule will not be a part of the case, but a way of interpreting and linking cases. Jonsen treats this problem by distinguishing descriptive elements in a case from moral maxims that inform judgment about the case: "These maxims provide the 'morals' of the story. For most cases of interest, there are several morals, because several maxims seem to conflict. The work of casuistry is to determine *which maxim* should *rule the case* and to what extent."[107] So understood, casuistry presupposes rather than defeats the claim that principles (or maxims or rules) are essential moral elements. The principles are held prior to the decision, and then are selected and weighed in the circumstances.

Further, just as Kant and many other philosophers have a problem with conflicting principles, so casuists have a problem with conflicting analogies and judgments. Cases that are amenable to many competing judgments, including the choice of analogies, are common in ethics. In a given sequence of events, discussants sometimes even see different *cases*. It is not enough to be told that cases point beyond themselves and evolve into generalizations. Perhaps cases will evolve in the wrong way because they were wrongly treated from the outset. Casuists have no clear methodological resource to prevent a biased development of cases and a neglect of relevant features of cases.

This problem leads to questions regarding the justificatory power of casuistry. How does *justification* occur? The casuists' answer rests on social convention and the patterns of judgment traced through their methods. But given the many different types of appeal that might be made (analogies, generalizations, character judgments, etc.), there apparently can be many different "right" answers on a single occasion. Without some stable framework of general norms, there is no control on judgment and no way to prevent prejudiced or poorly formulated social conventions.

Do case judgments have epistemic priority? Jonsen and Toulmin argue for the priority of cases and case judgments in moral knowledge, as does Brody for moral intuition and intuitive perceptiveness. Such claims are sometimes followed in casuistic arguments by a qualifying statement to the effect that particular moral judgments amend and augment general norms, but do not displace them. This qualification, however, seems to undermine rather than to qualify the priority thesis, and in any event is not the best model. If, as we argued in Chapter 1, a relationship of mutual adjustment exists between general norms and particular circumstances, neither the general nor the particular should be

granted an *order of priority*. The justification of beliefs moves from generalizations to cases *and* from cases to generalizations.

An overreaction to principles. An ambivalence about principles in the casuistry of Jonsen and Toulmin has never been consistently handled, raising questions about both the clarity and the coherence of the approach. On the one hand, they suggest a limited, conditional role for principles, and Jonsen explicitly says, "This casuistic analysis does not deny the relevance of principle and theory."[108] On the other hand, they eschew the use of principles such as those in this book, denounce firm and firmly held principles as tyrannical, and call the use of principles a "moralistic" use that is "not a serious ethical analysis."[109] This ambivalence mars their analysis and promotes an unnecessary rejection of principles.

Consider again Jonsen and Toulmin's example of the National Commission for the Protection of Human Subjects. Their constructive account of the commission's method of deliberation is unobjectionable. Commissioners reported that they were impressed by a history in medicine of cases of diagnostic, therapeutic, and preventive measures that formed a basis for their conclusions. However, the commissioners also reported a locus of certitude in moral principles. The transcripts of the commission's deliberations show a constant back-and-forth movement from principle to case, and from case to principle. Cases or examples favorable to one point of view were brought forward, and counterexamples to those cases were then advanced by a second commissioner against the examples and claims of the first. Principles were invoked to justify the choice and use of both examples and counterexamples. On many occasions a suggestion was made that a principle needed modification in light of a case, or an argument was offered that a case judgment was irrelevant or immoral in light of the commitments of a principle.[110] The commission's deliberations and conclusions are best understood, then, as examples of dialectical reasoning in which principles become interpreted, modified, and specified in context by the force of examples and counterexamples drawn from real-life cases.[111]

Jonsen and Toulmin appear to confuse the lack of a practical need for *theory* with the lack of a practical need for *principles*, as well as certitude about principles with certitude about theory. We believe the commission, the general public, and the mainstream of moral philosophy find a locus of certitude in the principles we present in this book, which do not sharply differ from the principles accepted by the commissioners. We agree that in practical deliberation we often do have more certitude about particular cases and conclusions (as well as maxims) than we do about various moral *theories*. But Jonsen and Toulmin suggest a lack of certitude surrounding *principles* (part of the common morality), when they should be pointing to the lack of certitude surrounding theories (which are not part of the common morality). Moreover, only deductivist theories seem damaged by their critical arguments. Theories with principles of prima facie obligation do not seem similarly damaged.

Mill argued that utilitarianism and other general theories can successfully face casuistical reservations about theory and general standards:

> There is no ethical creed which does not temper the rigidity of its laws, by giving a certain latitude, under the moral responsibility of the agent, for accommodation to peculiarities of circumstances; and under every creed, at the opening thus made, self-deception and dishonest casuistry get in. There exists no moral system under which there do not arise unequivocal cases of conflicting obligation. . . . They are overcome practically with greater or with less success according to the intellect and virtue of the individual; but it can hardly be pretended that any one will be the less qualified for dealing with them, from possessing an ultimate [general] standard to which conflicting rights and obligations can be referred.[112]

Mill rightly believes that the person skilled in case judgments will be aided by, not hindered by, general moral standards. We agree that every significant normative ethical theory must face the limits of its principles and rules, the necessity of moral judgment, the role of interpretation in particular cases, and the importance of circumstances. But, if so, what does casuistry add to our understanding of decisionmaking that these theories fail to provide? Does casuistry deserve to score higher on the criterion of *practicability* than the other theories we have studied? If not, the failure would be particularly devastating, because the major goal of recent casuistry has been to reach a practicable method.

A Constructive Evaluation of Casuistry

Today's casuists have resourcefully reminded us of the importance of analogical reasoning, paradigm cases, and practical judgment. Biomedical ethics, like ethical theory, has unduly minimized this avenue to moral knowledge. Casuists also have rightly pointed out that generalizations are often best learned, accommodated, and implemented by using cases, case discussion, and case methods. These insights can be utilized by connecting them to an appropriate set of concepts, principles, and theories that control the selection and analysis of cases. Biomedical ethics has long been driven by two kinds of analysis: case study and ethical theory. Cases such as *Quinlan*, *Bouvia*, and *Tarasoff* are discussed across the literature of the field, form a shared resource, and become integral to the way we think and draw conclusions. They profoundly influence our standards of fairness, negligence, paternalism, and the like.

Finally, a proper account of moral judgment is critical for biomedical ethics, which cannot flourish without a link between theory, principles, and decisionmaking. Sensitivity to context and individual differences is essential for a discerning use of principles. Casuistry would be notable if for no other reason than its long history of attempting to deal with this problem.

Principle-Based, Common-Morality Theories[113]

We will now turn attention to theories that both find their source in the common morality and use principles as their structural basis. A common-morality theory takes its basic premises directly from the morality shared in common by the members of a society—that is, unphilosophical common sense and tradition. Such a theory need not be principle-based, but we treat these two types of theories together in order to develop the tradition of ethics in which our account should be situated. This section, then, is best understood overall as a statement of the type of ethical theory that we accept and utilize in subsequent chapters.

Principle-based theories share with utilitarian and Kantian theories an emphasis on principles of *obligation*, but these theories share little else. Two main differences distinguish them. First, utilitarianism and Kantianism are *monistic* theories. One supreme, absolute principle supports all other action-guides in the system. Common-morality theories, as we here stipulatively define them, are *pluralistic*.[114] Two or more nonabsolute (prima facie) principles form the general level of normative statement. Second, common-morality ethics relies heavily on ordinary shared moral beliefs for its content, rather than relying on pure reason, natural law, a special moral sense, and the like. The principles embedded in these shared moral beliefs are also usually accepted by rival ethical theories. Although not the most general principles in many normative theories, the principles are nonetheless accepted in most types of ethical theory. The four principles developed in Chapters 3–6 should be understood as principles of this description.

Any theory that eventuates in moral judgments that cannot be brought into reflective equilibrium with pretheoretical commonsense judgments will be considered seriously flawed. However, this is not to maintain *either* that (for reasons discussed in Chapter 1) a common-morality theory is merely a systematizing of commonsense judgments *or* that all *customary moralities* qualify as part of the *common morality*. An important function of the standards in the common morality (from which the principles we defend and their correlative rights are developed) is to provide a basis for the evaluation and criticism of actions in countries and communities whose customary moral viewpoints fail to acknowledge basic principles. A customary morality, then, is not synonymous with the common morality. The latter is a pretheoretic moral point of view that transcends merely local customs and attitudes. Analogous to beliefs in the universality of basic human rights, the principles of the common morality are universal standards.

Our method in this book is to unite principle-based, common-morality ethics with the coherence model of justification delineated in Chapter 1. This strategy allows us to rely on the authority of the indispensable principles in the common morality, while incorporating tools to refine and correct its weaknesses and unclarities and to allow for additional specification. Because our strategy accepts the goal of reflective equilibrium and, in part, *constructs* principles and rules from considered judgments in the common morality, while also *specifying* principles and rules, we will not end with the identical content with which we began.

We can again illustrate this type of theory by reference to the case of the daughter who needs a kidney transplant. Unlike utilitarian and Kantian strategies, common-morality theories have no overarching principle to justify obligations or to adjudicate conflicts. Judgment requires interpreting as well as weighing and balancing moral norms to determine whether to respect the father's refusal to donate or to prod him into donating, whether to protect confidentiality or to lie, and whether to involve or to exclude the girl's siblings. When the father refuses to donate his kidney, the principle of respect for autonomy and related rules of privacy and liberty require that his choice not be forcibly overridden. These principles and rules are not absolute, but they do have sufficient weight in these circumstances to preclude forcible or coercive intervention to try to save the daughter. However, the physician has the right and perhaps even the responsibility to try to persuade the father to donate, at least by explaining and balancing probable benefits to the daughter and risks to the father. With a sufficiently high probability of successful transplantation and a sufficiently low risk to himself, the father may well have an obligation to donate, based on parental responsibilities. At some level of probable benefits and risks, the father's decision not to donate falls short of the ideal of parental love, and is therefore morally deficient, although no moral grounds warrant compelling him to donate.

Regarding the father's request to the physician to tell the family that he is not histocompatible, his predicted outcome of wrecking the family is a morally relevant consideration. But theories based on the common morality (like many other theories) would inquire whether alternatives short of lying or nondisclosure, such as counseling, could also prevent this outcome. A conflict is also present in this case between rules of truth-telling and confidentiality. Although direct lying is not always wrong from the perspective of prima facie principles, it requires principled justification. For example, lying is sometimes justified to shield a vulnerable teenager who does not want to donate a kidney to a sibling. The nature and defense of such judgments require us to examine further how principles function in common-morality theories.

The Common Morality as Primary Source

As a rough generalization, what Henry Sidgwick called the commonsense morality (morality's core principles and assorted rules of veracity, fidelity, and the like) is the source of the initial moral content for this type of theory. Ethical theory augments this spare content by a method (1) to clarify and interpret the content, (2) to make the various strands coherent, and (3) to further specify and balance the requirements of norms (as discussed in Chapter 1).

Consider why the common morality should play an essential role in ethical theory. If we could be confident that some abstract moral theory was a better source for codes and policies than the common morality, we could work constructively on practical and policy questions by progressive specification of the norms in that theory. But fully analyzed norms in ethical theories are invariably more contestable than the norms in the common morality. We cannot reasonably expect that a contested moral theory will be better for practical decisionmaking and policy development than the morality that serves as our common denominator. Far more social consensus exists about principles and rules drawn from the common morality (for example, our four principles) than about theories. This is not surprising, given the central social role of the common morality and the fact that its principles are, at least in schematic form, usually embraced in some form by all major theories. Theories are rivals over matters of justification, rationality, and method, but they often converge on mid-level principles (see pp. 109–111 below).

Common-morality ethics does not preclude the possibility of reform, which often occurs through interpretation, specification, and balancing. We earlier noted John Mackie's observation that interpretation and innovation are almost always carried out by appeal to justifications *within* rather than *beyond* norms already shared in the community. For example, if our policies on AIDS are so uncompassionate that we need to alter our conception of how therapeutic drugs are brought to the market, purchased, and distributed, this reevaluation will invoke available conceptions of compassion, fair funding, and distribution, rather than totally new principles of justice. Moreover, social agreements, traditions, and norms are inherently indeterminate, thereby failing to adequately anticipate the full range of moral problems and solutions. Interpretation and specification of norms, reconstruction of traditional beliefs, balancing different values, and negotiation are essential. This approach to construction in theory invites evolutionary change while insisting that the common morality provides the starting point and the constraining framework.

Two Examples of Principle-Based Theories

Commonsense convictions played only a minor role in ethical theory prior to the eighteenth century, when philosophers such as Francis Hutcheson, Jean-

Jacques Rousseau, and Joseph Butler argued that a native moral sense or an intuitive conscience possessed by all persons is far more important in the moral life than the more complicated systems of philosophers. Their moral psychology did not survive, but their commonsense emphasis did, and Hume, Kant, Hegel, and other leading moral theorists were deeply affected by it. Two twentieth-century writers in ethical theory will serve here to illustrate how a principle-based, common-morality theory is still alive and well.

Frankena's theory. An elegant and simple example of a common-morality theory that resembles ours is William Frankena's version of Hume's postulate that the two major "principles of morals" are beneficence and justice. Frankena appeals to what Bishop Butler called "the moral institution of life," together with what Frankena calls "the moral point of view," meaning a dispassionate attitude of sympathy in which moral decisions are reached by appeal to principled good reasons. For Frankena, the principle of beneficence (presented below on pp. 190–192, 260ff) resembles, but is not identical to, the utilitarian demand that we maximize good over evil, whereas the principle of justice (primarily an egalitarian principle) guides "our distribution of good and evil" independently of judgments about maximizing and balancing good outcomes. Frankena's theory comprises these two general principles, together with an argument that they capture the essence of the moral point of view.[115]

Ross's theory. A second example is the ethics of W. D. Ross, who has had a particularly imposing influence on twentieth-century ethical theory, and more influence on the present authors than any recent writer in ethical theory. He is best known for his intuitionism and his scholarship on Aristotle, but we will largely ignore these dimensions of his work. Ross's starting point is Aristotelian. The moral convictions of thoughtful persons are "the data of ethics just as sense-perceptions are the data of a natural science. Just as some of the latter have to be rejected as illusory, so have some of the former."[116] The "plain" person is, for Ross, the beginning rather than the end of the matter. Using this data base of ordinary standards, Ross thinks *acts* are properly categorized as right and wrong, whereas *motivation* and character are good and bad. This allows him to say that a right act can be done from a bad motive and that a good motive may eventuate in a wrong act.

Ross defends several basic and irreducible moral principles that express prima facie obligations. For example, promises create obligations of fidelity, wrongful actions and debts create obligations of reparation, and the generous services or gifts of others create obligations of gratitude. In addition to fidelity, reparation, and gratitude, Ross lists obligations of self-improvement, justice, beneficence, and nonmaleficence.[117] He holds that the principle of nonmaleficence (noninfliction of harm) takes precedence over the principle of beneficence (production of benefit) when the two come into conflict, but he assigns no

priorities among the other principles. This list of obligations is not grounded in any overarching principle.

In a noteworthy methodological statement, Ross maintains that principles are "recognized by intuitive induction as being implied in the judgments already passed on particular acts."[118] His studies of Greek philosophy also led him to distinguish knowledge from opinion. We know principles in the same way the plain person knows the main lines of moral obligation. Here we have *knowledge*, not *opinion*. However, when two or more obligations conflict and balancing, overriding, and judgment are necessary, Ross says we must examine the situation carefully until we form a "considered opinion (it is never more)" that one obligation is more incumbent in the circumstances than any other.[119] These judgments are about the *weight* of principles. They are not judgments that straightforwardly *apply* principles.

The Centrality of Principles and Rules

We can now develop the perspectives and assumptions in this book that make it a form of common-morality ethics.

The source of the principles. To say that principles have their origins in the common morality is not to suggest that the final form in which they greet a reader of this book is identical to their appearance in the common morality. Conceptual clarification and methods to introduce coherence are needed to give shape and substance to our moral commitments, much as grammarians, lexicographers, and stylists investigate the nature of our commitments in using words, punctuation, forms of citation, and the like. If unacceptable content is discovered in formulations of principles (for example, if a vigorous strong paternalism in clinical medicine is uncovered) or if incoherence is located, an attempt is made to find acceptable content and achieve coherence. This is work *in* ethical theory, even if its product should not be spoken of as *an* ethical theory. The objective is to give each principle a precise, plausible, thorough, and independent statement, without presupposing that our familiar ways of formulating principles are necessarily the best or the most coherent ways. After the principles are so formulated, they will still have to be further interpreted, specified, and balanced to produce an ethics for biomedicine. This is the heart of our strategy.

The prima facie and specifiable nature of the principles. Like Ross, we construe principles as prima facie binding. Some theories recognize rules, but treat them as expendable rules of thumb that summarize past experience by expressing better and worse ways to handle recurrent problems. Other theories contain absolute principles. Still other theories give a hierarchical (or lexical) ordering

to moral norms. We reject all three interpretations as inadequate to capture the nature of moral norms and moral reasoning. Rules of thumb permit too much discretion, as if principles or rules were not binding; absolute principles and rules disallow all discretion for moral agents and also encounter unresolvable moral conflicts; and a hierarchy of rules and principles suffers from damaging counterexamples whose force depends on our reservoir of considered judgments. (Unlike Ross, we assign no form of priority weighing or hierarchical ranking to our principles.)

By contrast, we treat principles as both prima facie binding and subject to revision. So understood, a prima facie principle is a normative guideline stating conditions of the permissibility, obligatoriness, rightness, or wrongness of actions that fall within the scope of the principle. The latitude to balance principles in cases of conflict leaves room for compromise, mediation, and negotiation. The account is thereby rescued from the charge that principles cannot be compromised and so become tyrannical. In stubborn cases of conflict there may be no single right action, because two or more morally acceptable actions are unavoidably in conflict and yet have equal weight in the circumstances. Here we can give good but not decisive reasons for more than one action.

For instance, although murder is absolutely prohibited because of the normative content in the word *murder*, it is not plausible to hold that killing is absolutely prohibited. Killing persons is *prima facie* wrong, but killing to prevent a person's further extreme pain or suffering is not wrong in every circumstance. Killing may be the only way to meet some obligations, even though it is prima facie wrong (see Chapter 4, pp. 219–241.) However, when a prima facie obligation is outweighed or overridden, it does not simply disappear or evaporate. It leaves what Nozick calls "moral traces,"[120] which should be reflected in the agent's attitudes and actions.

A disadvantage of this account, some say, is that it moves relentlessly to the paradoxical conclusion that, as Hume put it, "the principles upon which men reason in morals are always the same; though the conclusions which they draw are often very different."[121] True, a relativity of judgment is inevitable, but a relativity of the principles embedded in the common morality is not. When people reach different conclusions, their moral judgments are still subject to justification by good reasons. They are not purely arbitrary or subjective judgments. A judgment can be proposed for consideration on any basis a person chooses—random selection, emotional reaction, mystical intuition, etc.—but to propose is not to justify, and one part of justification is to test judgments and norms by their coherence with the other norms in the moral life.

We conclude that although flexibility and diversity in judgment are ineliminable, judgment generally should be constrained by the demands of moral justification, which typically involves appeal to principles. Our presentation of the principles—together with arguments to show the coherence of these principles

with other aspects of the moral life, such as the moral emotions, virtues, and rights—*constitutes* the theory in the present volume. This web of norms and arguments *is* the theory. There is no single unifying principle or concept, no description of the highest good, and the like.

A Reply to Some Criticisms

Some commentators have criticized our account as a mere "mantra of principles," meaning that the principles often function like a ritual incantation of norms repeated with little reflection. H. Danner Clouser and Bernard Gert have so argued in an attack on "principlism," a term they use to designate all theories composed of a plural body of potentially conflicting prima facie principles—principally our account and Frankena's. They accuse us of the following defects in theory:[122] (1) The "principles" are little more than names, checklists, or headings for values worth remembering, leaving principles without deep moral substance or capacity to guide action. (2) Principle-analyses fail to provide a unified theory of justification or a general theory that ties the principles together as a systematic, coherent, and comprehensive body of guidelines, with the consequence that the alleged action-guides are ad hoc constructions lacking systematic order. (3) The prima facie principles (and other action guides in the framework) often conflict, and the underlying account is too indeterminate to provide a decision procedure to adjudicate those conflicts.

We do not deny that these problems are worthy of sustained reflection. We reject, however, certain assumptions that Clouser and Gert make, especially their requirement that there be "a single clear, coherent, and comprehensive decision procedure for arriving at answers."[123] We are skeptical of this enterprise, even as a model for ethical theory, for reasons presented in Chapter 1 (see pp. 13–37). Regarding their *first* criticism, that our principles are checklists or headings without deep moral substance, we agree that principles order, classify, and group moral norms that need additional content and specificity. Until the principles are *interpreted and analyzed* (as they are in every first section of Chapters 3–6) and *specified and connected to other norms* (as they are in later sections of these chapters), it is unreasonable to expect much more than a classification scheme that organizes the normative content.[124]

Regarding the *second* criticism,[125] that our principle-analysis fails to provide a systematic theory, we see the point but view it as irrelevant. We have not attempted a general ethical theory and do not claim that our principles mimic, are analogous to, or substitute for the foundational principles in leading classical theories such as utilitarianism (with its principle of utility) and Kantianism (with its categorical imperative). We have expressed a constrained skepticism about this foundationalism and are doubtful that such a unified foundation for ethics is discoverable. As we have acknowledged, even the core principles in our account are so scant that they cannot provide an adequate basis for deducing most of what we can justifiably claim to know in the moral life.

Regarding the *third* criticism, that principles compete in ways our account cannot resolve, we acknowledge that conflicts among principles cannot be resolved a priori. No system of guidelines could reasonably anticipate the full range of conflicts, and the point of our discussion of dilemmas was to indicate circumstances in which principles (and other commitments) pull us in different directions. No one escapes this problem in living the moral life. It does not follow that these principles are inconsistent or that we encounter incompatible moral commitments by accepting these principles. It is a virtue of our theory that it requires specification, and a defect in Clouser and Gert's account that it purports through its rules to escape the need for specification. Only a theory that could put enough content in its norms to escape conflicts and dilemmas in all contexts could live up to the Clouser-Gert demand, and no theory has come close to doing so. It is therefore essential to leave room for interpretation, specification, and balancing of principles and rules in the face of recurrent and recalcitrant conflicts.

Here experience and sound judgment are indispensable allies. It is insupportably optimistic, for several reasons, to suppose we have attained or will attain a fully specified system of norms for health care ethics. Thomas Nagel has forcefully argued that an unconnected heap of obligations and values is an ineradicable feature of morality, and Ross rightly argued that his Kantian and utilitarian critics forced an "architectonic" of "hastily reached simplicity" on ethics. Whereas critics of Ross's account (and ours) rely on an ideal of systematic unity, we see disunity, conflict, and moral ambiguity as pervasive aspects of the moral life. Untidiness, complexity, and conflict are unfortunate features of communal living, but a theory of morality cannot be faulted for a realistic appraisal of them.

Persons typically lack a complete understanding of the full range of commitments they make in accepting a principle, because of its indeterminateness and our inability to specify it all the way down to the concrete cases. We come to understand principles, and what they exclude and include, by making judgments in particular circumstances. For example, many of us likely have a poor idea of the demands of the principle of respect for autonomy in an institution serving persons with serious mental disabilities. But if we were in that environment on a daily basis, we probably would develop a better set of ideas. The judgments that we make achieve an increased specificity, which then loops back and in some cases forces a revised gloss on the norms in our moral framework.

This growth of moral understanding is to be encouraged. Only a faulty conception of the nature and interpretation of principles would lead to the conclusion that principles have no integral role in moral reasoning in concrete circum-

stances. The more accurate estimate is that principles point us in the right direction, but we then typically encounter a host of other considerations that must be accommodated, such as institutional practices, limited resources, judgments about acceptable risk, religious beliefs, and personal projects and aspirations.

Critical Evaluation

Although we accept a version of principle-based, common-morality ethics, we acknowledge that this approach has unresolved problems. Three are of interest for the present chapter on types of theory.

Specification and judgment. Do principles, when specified for behavior, enable us to reach practical judgments, or are they either too indeterminate or too determinate to eventuate in judgments? Again we confront a problem of practicability. We must be careful both to specify in order to escape abstractness and not to *overspecify* a principle or rule, because it then becomes too rigid and insensitive to circumstances. We have argued that the best course of action is to accept both abstract principles and a method for specifying those principles, in order that they can be appropriately implemented for specific circumstances. But can this goal be achieved in practice? For example, if in specifying the principle of respect for autonomy and rules of truth-telling in medicine, the resultant rules require too little education of and conversation with patients in some cases and too much in other cases (where the health care system will not bear the time and cost of the requirements), then the specification is inadequate. Many specified principles and rules will encounter this problem of too-little and too-much for some contexts, which is one reason why balancing and judgment are as consequential as specification. But without tighter controls on permissible balancing than common-morality theories propose, critics charge that too much room is left for judgments that are unprincipled and yet sanctioned or permitted by the theory. Can the conditions on balancing presented in Chapter 1 reduce intuition to an acceptable level? Can the constraints of coherence be tightened to adequately respond to these concerns? In Chapter 8, we will argue that these problems are actually more complicated, requiring acceptance of important parts of character ethics and a distinction between principles of obligation and the judgments that are guided by those principles.

Can the common morality be rendered coherent? We have linked a coherence theory of justification to a common-morality theory, but can the common morality be made coherent? If one argues (as we did above, citing Nagel) that an unconnected heap of obligations and values comprises the common morality, is there any hope of rendering the heap coherent, short of so radically reconstructing the norms that they become only a distant cousin of the common

morality? Is the goal of coherence more an article of faith than a demonstrable achievement?

Consider, for example, the use of animals as research subjects. Across society we disagree regarding the "considered judgments" that might form our starting place, and it is doubtful that the common morality can be tapped to find shared considered judgments. On the one hand, one might argue that we need to search for a wider body of beliefs that can be brought into reflective equilibrium. On the other hand, we might simply specify our personal considered judgments. Either way, no clear starting point and body of beliefs exist to be rendered coherent, mainly because of a lack of agreement about the status of animals and our obligations, if any, to them. Many other controversies similarly invite disagreement over appropriate *initial* beliefs.

Is there a theory to be constructed? These problems of coherence lead straight to some connected problems about theory. The language of a "common-morality *theory*" suggests either that a theory underlies the common morality or that a theory can be philosophically constructed from it. Is there good reason to believe that a theory (not merely an unconnected group of coherent principles and rules) is possible? Perhaps mid-level principles, polished analyses of the moral virtues, and coherent statements of transnational human rights are all that should be attempted, rather than a theory that conforms to the criteria delineated at the beginning of this chapter. Perhaps "moral theory" has been so diluted in meaning in "common-morality theories" that the goal of a theory should be abandoned altogether, in favor of a more modest goal, such as "moral reflection and construction." A related problem is that attempts to bring the common morality into greater coherence risk decreasing rather than increasing moral agreement in society. That is, a theory can introduce claims that generate disagreements not found in the initial considered judgments; or, as we have often seen in the history of ethics, the theory may turn out to be less clear and reliable for practical decisionmaking than the common morality.

In part these problems turn on one's expectations for a "theory." Clouser and Gert expect a strong measure of unity and systematic connection among rules, a clear pattern of justification, and a practical decision procedure that flows from a theory, whereas Annette Baier is skeptical of each of these conditions, and even of the language of "theory."[126] We need not here debate the theory of theories, but we do need to return to the theme of convergence among ethical theories.

Convergence Across Theories

Whenever several competing theories or systems of belief are available, we seek out the best and affiliate with it, while rejecting the others. However, affiliation with one type of theory is not always the best strategy in either

general ethics or biomedical ethics. If the two authors of this book were forced to rank the types of theory examined in this chapter other than common-morality theory, we would differ. We have reached different estimates after testing available theories under the criteria established early in this chapter. But for both of us, the most satisfactory theory—if we could find *one* to substitute for a common-morality theory—would be only slightly preferable, and no theory would fully satisfy all the criteria.

Differences between types of theory are exaggerated if they are presented as warring armies locked in combat. Many different theories lead to similar action-guides and to similar estimates of the role of character in ethics. It is possible from several of these standpoints to defend roughly the same principles, obligations, rights, responsibilities, and virtues. For example, although utilitarianism is often depicted both as starkly different from and as hostile to the other theories, when utilitarian Richard Brandt states his view, it is strikingly reminiscent at the level of principle and obligation to Ross's estimate, which we have seen to be sharply critical of utilitarianism:

> [The best code] would contain rules giving directions for recurrent situations which involve conflicts of human interests. Presumably, then, it would contain rules rather similar to W. D. Ross's list of prima facie obligations: rules about the keeping of promises and contracts, rules about debts of gratitude such as we may owe to our parents, and, of course, rules about not injuring other persons and about promoting the welfare of others where this does not work a comparable hardship on us.[127]

That Brandt appeals to utility and Ross to intuitive induction to justify similar sets of rules is a significant difference at the level of moral justification, and the two authors might interpret and specify their rules differently. Yet, they exhibit only trivial differences in their lists of primary obligations. This convergence is not restricted to Brandt and Ross. It is common in normative theories that provide frameworks of principles and rules. Such agreement springs in part from an initial shared data base, namely, the norms of the common morality.

This convergence offers encouragement to practical ethics, although in itself it does not resolve either theoretical differences or practical problems. Convergence as well as consensus about principles among a group of persons is common in assessing *cases* and framing *policies*, even when deep theoretical differences divide the group. Agreement may similarly be reached regarding precedent cases. This is not to deny that theoretical differences do sometimes eventuate in practical disagreements, in different policies, and in unresolvable dilemmas. For example, utilitarians tend to support various types of research involving human subjects because of the potential benefits the research offers for future patients. Many nonutilitarians tend to be skeptical of some of this research on grounds of its actual or potential violation of individual rights. But utilitarians and their theoretical opponents also often cross over these lines of demarcation and agree that any adequate ethical approach to research involving human subjects must include some of the constraints and considerations that

have been highlighted by proponents of both types of theory (and that are now embedded in the major codes and regulations of research involving human subjects).

Reasons exist, then, for holding that distinctions between types of theory are not as significant for *practical* ethics as has sometimes been proclaimed. It is a mistake to suppose that a series of continental divides separates moral theorists into distinct and hostile groups who reach different practical conclusions and fail to converge on principles. We should not overlook the fact that some theories are closer in substantive principles and rules to supposedly rival theories than they are even to some theories of their own "type."

Conclusion

Contemporary biomedical ethics incorporates theoretical conflicts of considerable complexity, and the diverse theories explored in this chapter help us see why. Competition exists among the various normative theories, and in addition we find a body of competing conceptions as to how such theories should be related to biomedical practice. Thus, persons who agree on a particular type of ethical theory may still find themselves in sharp disagreement regarding how to relate their theory to the treatment of particular moral problems.

Nonetheless, we stand to learn from all of these theories. Where one theory is weak in accounting for some part of the moral life, another is often strong. Although each type of theory clashes at some point with deep moral convictions, each also articulates norms that we are reluctant to relinquish. Each of the theories discussed in this chapter has led to the development and rejection of prominent hypotheses in moral theory. Although we have described our approach as principle-based, we reject the assumption that one must defend a single type of theory that is solely principle-based, virtue-based, rights-based, case-based, and so forth. In moral reasoning we often blend appeals to principles, rules, rights, virtues, passions, analogies, paradigms, parables, and interpretations. To assign priority to one of these factors as the key ingredient is a dubious project, as is the attempt to dispense with ethical theory altogether. The more general (principles, rules, theories, etc.) and the more particular (feelings, perceptions, case judgments, practices, parables, etc.) should be linked together in our moral thinking. We will have more to say about how these strands are mutually supportive in Chapters 7 and 8, after we develop our framework of principles in Chapters 3–6.

Notes

1. See Baruch Brody's definition of "pluralism" in *Life and Death Decision Making* (New York: Oxford University Press, 1988), p. 9. Our views on pluralism are

influenced by Thomas Nagel, "The Fragmentation of Value," in *Mortal Questions* (Cambridge: Cambridge University Press, 1979), pp. 128–37.

2. Our discussion in this edition has profited from Shelly Kagan, *The Limits of Morality* (Oxford: Clarendon Press, 1989), esp. pp. 11–15, and from criticisms by David DeGrazia.
3. For recent analysis of this utilitarian thesis, see Samuel Scheffler, *Consequentialism and its Critics* (Oxford: Clarendon Press, 1988).
4. Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. Burns & Hart (Oxford: Clarendon Press, 1970), pp. 11–14, 31, 34. John Stuart Mill, *Utilitarianism*, in vol. 10 of the *Collected Works of John Stuart Mill* (Toronto: University of Toronto Press, 1969), ch. 1, p. 207; ch. 2, pp. 210, 214; ch. 4, pp. 234–35.
5. A representative of the first list is G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903), pp. 90ff; a representative of the latter list is James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Clarendon Press, 1986), p. 67.
6. This case is based on Melvin D. Levine, Lee Scott, and William J. Curran, "Ethics Rounds in a Children's Medical Center: Evaluation of a Hospital-Based Program for Continuing Education in Medical Ethics," *Pediatrics* 60 (August 1977): 205.
7. Worthington Hooker, *Physician and Patient* (New York: Baker and Scribner, 1849), pp. 357ff, 375–81.
8. J. J. C. Smart, *An Outline of a System of Utilitarian Ethics* (Melbourne: University Press, 1961); and "Extreme and Restricted Utilitarianism," in *Contemporary Utilitarianism*, ed. Michael Bayles (Garden City, NY: Doubleday and Co., 1968), esp. pp. 104–7, 113–15.
9. Richard B. Brandt, "Toward a Credible Form of Utilitarianism," in *Contemporary Utilitarianism*, pp. 143–86, and in Brandt's *Morality, Utilitarianism, and Rights* (Cambridge: Cambridge University Press, 1992).
10. Alan Donagan, "Is There a Credible Form of Utilitarianism?" in *Contemporary Utilitarianism*, pp. 187–202.
11. A subtle argument to this conclusion is found in Kagan, *The Limits of Morality*, *passim*.
12. Williams, "A Critique of Utilitarianism," in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 116–17, and J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977), pp. 129, 133.
13. Milton Weinstein and William B. Stason, *Hypertension* (Cambridge, MA: Harvard University Press, 1977), and their articles in *New England Journal of Medicine* 296 (1977): 716–21, and *Hastings Center Report* 7 (October 1977): 24–29.
14. Amartya Sen, *On Ethics and Economics* (Oxford: Basil Blackwell, 1987), p. 75.
15. Kant sought to show that reason unaided can be and should be a proper motive to action. What we should do morally is determined by what we would do "if reason completely determined the will." *The Critique of Practical Reason*, trans. Lewis White Beck (New York: Macmillan, 1985), pp. 18–19. Ak. 20. "Ak." designates the page-reference system of the 22-volume Preussische Akademie edition conventionally cited in Kant scholarship.
16. Kant, *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck (Indianapolis, IN: Bobbs-Merrill Company, 1959), pp. 37–42; Ak. 421–24.
17. *Foundations*, p. 47; Ak. 429.
18. *Foundations*, pp. 51, 58–63; Ak. 432, 439–44.

19. *Foundations*, pp. 58; Ak. 439–40; and *Critique of Practical Reason*, p. 33; Ak. 33.
20. Alan Donagan, *The Theory of Morality* (Chicago: University of Chicago Press, 1977), pp. 63–66.
21. See *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), pp. 3–4, 26–31. For Rawls's more technical interests in and development of Kant, see his "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford, CA: Stanford University Press, 1989), pp. 81–113.
22. *A Theory of Justice*, pp. 102, 137, 252–55. In § 40 of *A Theory of Justice*, Rawls presents his "Kantian Interpretation of Justice as Fairness," as "based upon Kant's notion of autonomy."
23. Sidgwick, *The Methods of Ethics*, 7th Ed. (Indianapolis, IN: Hackett Publishing Co., 1981), p. 516.
24. Rawls, *A Theory of Justice*, pp. 252, 256, 515–19. See also, "A Kantian Conception of Equality," *Cambridge Review* (February 1975): 97ff.
25. Rawls, "The Priority of Right and Ideas of the Good," *Philosophy & Public Affairs* 17 (1988): 252, and "Justice as Fairness: Political not Metaphysical," *Philosophy & Public Affairs* 14 (1985): 223–51, esp. 224–25.
26. "The Doctrine of Virtue," Part II of *The Metaphysic of Morals*, trans. Mary J. Gregor (Philadelphia: University of Pennsylvania Press, 1964), p. 85, Ak, 421–22; See also *Lectures on Ethics*, ed. Louis Infield (New York: Harper and Row, 1963), pp. 150–54.
27. G. W. F. Hegel, *Philosophy of Right*, trans. T. M. Knox (Oxford: Clarendon Press, 1942), pp. 89–90, 106–7.
28. For two influential anthologies of recent work, see *Midwest Studies in Philosophy Volume XIII—Ethical Theory: Character and Virtue*, ed. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (Notre Dame, IN: University of Notre Dame Press, 1988); and *Identity, Character, and Morality*, ed. Owen Flanagan and Amélie Oksenberg Rorty (Cambridge, MA: MIT Press, 1990). For two very different treatments of the Aristotelian perspective, see Nancy Sherman, *The Fabric of Character: Aristotle's Theory of Virtue* (Oxford: Clarendon Press, 1989); Alasdair MacIntyre, *After Virtue*, 2d Ed. (Notre Dame, IN: University of Notre Dame Press, 1984).
29. This is not the broadest sense, inasmuch as machines, tools, horses, and the like are often said to have virtues. Some writers more tightly restrict the meaning of *virtue* than we do. For example, Aristotle required that virtue involve habituation rather than a natural character trait. *Nicomachean Ethics*, trans. Terence Irwin (Indianapolis, IN: Hackett Publishing Co., 1985), 1103a18–19. Thomas Aquinas (relying on a formulation by Peter Lombard) additionally held that virtue is a good quality of mind by which we live rightly and therefore *cannot be put to bad use*. *Treatise on the Virtues* (from *Summa Theologiae*, I–II), Question 55, Arts. 3–4, pp. 54–55.
30. This definition is the primary use reported in the *O.E.D.* It is defended by Alan Gewirth, "Rights and Virtues," *Review of Metaphysics* 38 (1985): 751, and R. B. Brandt, "The Structure of Virtue," in *Midwest Studies in Philosophy* 13 (1988): 76. Edmund Pincoffs presents a definition of virtue in terms of desirable dispositional qualities of persons, in *Quandaries and Virtues: Against Reductivism in Ethics* (Lawrence: University Press of Kansas, 1986), pp. 9, 73–100. We accepted a definition similar to these in the first two editions of this book, for which we were criticized by John Waide, "Virtues and Principles," *Philosophy and Phenomenological Research* 48 (1988): 455–72.

31. *Nicomachean Ethics*, bk. II, 1105[a]17–33, 1106[b]21–23; cf. bk. VI, 1144[a]14–20 (trans. Irwin).
32. See Philippa Foot, *Virtues and Vices* (Oxford: Basil Blackwell, 1978), Rodger Beehler, *Moral Life* (Oxford: Basil Blackwell, 1978), Gregory Trianosky, "Supererogation, Wrongdoing, and Vice," *Journal of Philosophy* 83 (1986): 26–40.
33. See Michael Stocker, "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy* 73 (1976): 453–66. One might try to amend Kant to overcome this objection. See Kurt Baier, "Radical Virtue Ethics," *Midwest Studies in Philosophy* 13 (1988): 130–31.
34. H. K. Beecher, "Ethics and Clinical Research," *New England Journal of Medicine* 274 (1966): 1354–60.
35. G. Pence, *Ethical Options in Medicine* (Oradell, NJ: Medical Economics Co., 1980), p. 177.
36. Thomas Keneally, *Schindler's List* (New York: Penguin Books, 1983), pp. 176–80.
37. This formulation is indebted to David Hume, *A Treatise of Human Nature*, 2d Ed., ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1978), p. 478.
38. See William K. Frankena, *Ethics*, 2d Ed. (Englewood Cliffs, NJ: Prentice-Hall, 1973), p. 65; and Kurt Baier, "Radical Virtue Ethics," pp. 133–34.
39. This schema has been adapted, with modifications, from Tom L. Beauchamp, *Philosophical Ethics*, 2d Ed. (New York: McGraw-Hill, 1991), ch. 6.
40. See David Solomon, "Internal Objections to Virtue Ethics," *Midwest Studies in Philosophy* 13 (1988): 439.
41. Hart, "Between Utility and Rights," in *Jurisprudence and Philosophy* (Oxford: Clarendon Press, 1983), p. 198. For debates about liberalism, see Nancy L. Rosenblum, ed. *Liberalism and the Moral Life* (Cambridge, MA: Harvard University Press, 1989).
42. Compare H. L. A. Hart, "Bentham on Legal Rights," in *Oxford Essays in Jurisprudence*, 2nd series. ed. A. W. B. Simpson (Oxford: Oxford University Press, 1973), pp. 171–98.
43. See Joel Feinberg, *Social Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1973), p. 67.
44. Dworkin, *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1977), pp. xi, 92, 191, and "Is there a Right to Pornography?" *Oxford Journal of Legal Studies* 1 (1981): 177–212.
45. See James Griffin, "Towards a Substantive Theory of Rights," in *Utility and Rights*, ed. R. G. Frey (Minneapolis: University of Minnesota Press, 1984), pp. 155–58.
46. See Judith Jarvis Thomson, *The Realm of Rights* (Cambridge, MA: Harvard University Press, 1990), pp. 122ff.
47. See Feinberg, *Social Philosophy*, p. 59; and Eric Mack, ed., *Positive and Negative Duties* (New Orleans: Tulane University Press, 1985).
48. The first decade of decisions began with *Roe v. Wade* 410 U.S. 113 (1973) and ran through *City of Akron v. Akron Center for Reproductive Health* (June 1983). Decisions of major importance pertaining to indigency and funding were *Maher v. Roe*, 432 U.S. 464 (1977) and *Harris v. McRae*, 448 U.S. 297 (1980). In *Planned Parenthood v. Casey* (June 28, 1992), the U.S. Supreme Court further upheld the pregnant woman's right to terminate her pregnancy within limits, while abolishing the trimester framework. It recognized the state's interest in fetal life from the beginning of the pregnancy and allowed states to institute requirements that do not impose an undue burden on the pregnant woman's decisions and actions.

49. See A. I. Melden, *Rights and Right Conduct* (Oxford: Basil Blackwell, 1959).
50. See David Braybrooke, "The Firm but Untidy Correlativity of Rights and Obligations," *Canadian Journal of Philosophy* 1 (1972): 351–63.
51. Mill, *Utilitarianism*, in *Collected Works*, p. 247.
52. Ronald Dworkin argues that political morality is rights-based in *Taking Rights Seriously*, pp. 169–77, esp. 171. John Mackie has applied this thesis to morality generally in "Can There Be a Right-Based Moral Theory?" *Midwest Studies in Philosophy* 3 (1978), esp. p. 350.
53. Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. ix, 149–82.
54. Gewirth, "Why Rights are Indispensable," *Mind* 95 (1986): 333.
55. See Michael Sandel, "The Political Theory of the Procedural Republic," *Revue de métaphysique et de morale* 93 (1988): 57–68, esp. 64–67; Sandel, "Democrats and Community," *The New Republic*, February 22, 1988: 20–23; Alasdair MacIntyre, *After Virtue*, pp. 235–37; Michael Walzer, "The Communitarian Critique of Liberalism," *Political Theory* 18 (1990): 6–23. See also Shlomo Avineri and Avner de-Shalit, eds., *Communitarianism and Individualism* (Oxford: Oxford University Press, 1992); David Rasmussen, ed., *Universalism vs. Communitarianism: Contemporary Debates in Ethics* (Cambridge, MA: MIT Press, 1990); and Donald L. Gelpi, ed., *Beyond Individualism: Toward a Retrieval of Moral Discourse in America* (Notre Dame, IN: University of Notre Dame Press, 1989).
56. See Sandel, *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1982), pp. 15–17.
57. Ibid., p. 172; cf. p. 179. See similar statements in MacIntyre, *After Virtue*, pp. 203–6.
58. Sandel, "Introduction," in *Liberalism and Its Critics*, ed. Sandel (New York: New York University Press, 1984), p. 6, and "Morality and the Liberal Ideal," *The New Republic* (May 7, 1984), pp. 15–17; MacIntyre, *After Virtue*, ch. 1.
59. Sandel, *Liberalism and the Limits of Justice*, 15–23, 84–87, 92–94, 139–51; MacIntyre, *Whose Justice? Which Rationality?* (Notre Dame, IN: University of Notre Dame Press, 1988), p. 10, and *After Virtue*, p. 206.
60. *After Virtue*, p. 53.
61. Mackie, *Ethics*, pp. 30, 36–37; see also 106–10, 120–24.
62. Alasdair MacIntyre, *After Virtue*, pp. 17, 187, 190–94.
63. "Morality and the Liberal Ideal," p. 17.
64. See James L. Nelson, "The Rights and Responsibilities of Potential Organ Donors: A Communitarian Approach," *Communitarian Position Paper* (Washington, DC: The Communitarian Network, 1992); James Muyskens, "Procurement and Allocation Policies," *The Mount Sinai Journal of Medicine* 56 (1989): 202–6.
65. For the wide range of issues in organ procurement, see James F. Childress, "Ethical Criteria for Procuring and Distributing Organs for Transplantation," *Journal of Health Politics, Policy and Law* 14 (1989): 87–113.
66. Callahan, *What Kind of Life* (New York: Simon and Schuster, 1990), ch. 4, esp. pp. 105–13, and *Setting Limits* (New York: Simon and Schuster, 1987), esp. pp. 106–14.

67. Ezekiel J. Emanuel, *The Ends of Human Life: Medical Ethics in a Liberal Polity* (Cambridge, MA: Harvard University Press, 1991). See also Troyen Brennan, *Just Doctoring: Medical Ethics in the Liberal State* (Berkeley: University of California Press, 1991).
68. See esp. Charles Taylor, *Philosophy and the Human Sciences*, Philosophical Papers, vol. 2 (Cambridge: Cambridge University Press, 1985), ch. 7, "Atomism."
69. See arguments to this conclusion in Will Kymlicka, "Liberalism and Communitarianism," *Canadian Journal of Philosophy* 18 (June 1988): 181–204, and "Liberal Individualism and Liberal Neutrality," *Ethics* 99 (July 1989): 883–905. Even in John Locke and Thomas Hobbes—the communitarians' arch-enemies (see MacIntyre, *After Virtue*, pp. 233–34)—there is a considerable emphasis on promoting the commonweal. Locke gives an elegant statement in *Two Treatises of Civil Government, Works* (London: C. and J. Rivington, 1824), 12th Ed., bk. 2, note 8, p. 357.
70. See Amy Gutmann, "Communitarian Critics of Liberalism," *Philosophy and Public Affairs* 14 (Summer 1985): 308–22.
71. Sandel interprets Rawls as creating this dilemma by his account of the priority of the right over the good. *Liberalism and the Limits of Justice*, pp. 1–10, 17–24, 168–72, and "Morality and the Liberal Ideal," pp. 16–17.
72. Joel Feinberg, *Harm to Self*, vol. 3 in *The Moral Limits of the Criminal Law* (New York: Oxford University Press, 1986), p. 47.
73. See MacIntyre, *Against Virtue*, pp. 67–68.
74. See Allen Buchanan, "Assessing the Communitarian Critique of Liberalism," *Ethics* 99 (July 1989): 852–82, esp. 862–65, and William A. Galston, *Liberal Purposes* (Cambridge: Cambridge University Press, 1991).
75. Thomson, *The Realm of Rights*, p. 223.
76. Carol Gilligan, *In a Different Voice* (Cambridge, MA: Harvard University Press, 1982), esp. p. 21. For these themes in her later work, see her "Mapping the Moral Domain: New Images of Self in Relationship," *Cross Currents* 39 (Spring 1989): 50–63.
77. Gilligan and many others deny that the two distinct voices correlate strictly with gender. See Gilligan and Susan Pollak, "The Vulnerable and Invulnerable Physician," in *Mapping the Moral Domain*, ed. C. Gilligan, J. Ward, and J. Taylor (Cambridge, MA: Harvard University Press, 1988), pp. 245–62.
78. See Gilligan and G. Wiggins, "The Origins of Morality in Early Childhood Relationships," in *The Emergence of Morality in Young Children*, ed. J. Kagan and S. Lamm (Chicago: University of Chicago Press, 1988). See also Sara Ruddick, *Maternal Thinking: Toward a Politics of Peace* (Boston: Beacon Press, 1989).
79. Baier, "What Do Women Want in a Moral Theory?" *Nous* 19 (March 1985): 53.
80. Ibid., pp. 53–56.
81. Cf. Baier, *Postures of the Mind* (Minneapolis: University of Minnesota Press, 1985), pp. 210–19.
82. Our formulation of these criticisms of liberalism is influenced by Alisa L. Carse, "The 'Voice of Care': Implications for Bioethical Education," *The Journal of Medicine and Philosophy* 16 (1991): 5–28, esp. 8–17.
83. Baier, "Trust and Antitrust," *Ethics* 96 (1986): 248.
84. Quill and Townsend, "Bad News: Delivery, Dialogue, and Dilemmas," *Archives of Internal Medicine* 151 (March 1991): 463–64.
85. See Nel Noddings, *Caring: A Feminine Approach to Ethics and Moral Education* (Berkeley: University of California Press, 1984).

86. See Nancy Sherman, *The Fabric of Character* (Oxford: Oxford University Press, 1989), pp. 13–55, and Martha Nussbaum, *Love's Knowledge* (Oxford: Oxford University Press, 1990).
87. See the Kantian arguments in Barbara Herman, "Integrity and Impartiality," *Monist* 66 (April 1983): 233–50, and Marcia Baron, "The Alleged Repugnance of Acting from Duty," *Journal of Philosophy* 81 (April 1984): 197–220.
88. For a recent attempt to incorporate caring into a framework that retains impartiality, see Jeffrey Blustein, *Care and Commitment: Taking the Personal Point of View* (New York: Oxford University Press, 1991).
89. Carse, "The 'Voice of Care,' " p. 17.
90. Susan Sherwin, *No Longer Patient: Feminist Ethics and Health Care* (Philadelphia: Temple University Press, 1992), pp. 49–50. See also Laura Purdy, "A Call to Heal Ethics," in Helen Bequaert Holmes and Purdy, eds., *Feminist Perspectives in Medical Ethics* (Bloomington: Indiana University Press, 1992), p. 10.
91. See Hilde L. Nelson, "Against Caring," Nel Noddings, "In Defense of Caring," and Toni M. Vezeau, "Caring: From Philosophical Concerns to Practice," in *The Journal of Clinical Ethics* 3 (Spring 1992): 8–20. Two recent anthologies provide a good entry point for these debates: Claudia Card, ed., *Feminist Ethics* (Lawrence: University Press of Kansas, 1991), and Eve Browning Cole and Susan Coultrap-McQuin, eds., *Explorations in Feminist Ethics: Theory and Practice* (Bloomington: Indiana University Press, 1992). See also Christine Overall, *Ethics and Human Reproduction: A Feminist Analysis* (Boston: Allen and Unwin, 1987).
92. Casuists have had little to say about the nature of a case, but see the analysis in Albert R. Jonsen, "Casuistry as Methodology in Clinical Ethics," *Theoretical Medicine* 12 (December 1991): 298.
93. For leading expositions, see Albert R. Jonsen and Stephen Toulmin, *The Abuse of Casuistry: A History of Moral Reasoning* (Berkeley: University of California Press, 1988) and Brody, *Life and Death Decision Making*.
94. The casuistical tradition was prominent in the seventeenth century when it came under a severe and enduring attack by Blaise Pascal in his *Provincial Letters* (18 letters published in 1656–57, under the pseudonym Louis de Montalte). Jonsen and Toulmin assess the relevance and fairness of Pascal's attack in *Abuse of Casuistry*, ch. 12, esp. pp. 243–49.
95. *An Encyclopedia of Religion*, ed. V. Ferm (New York: The Philosophical Library, 1945), p. 124.
96. Here are two candidates. Jeremy Bentham: "From utility then we may denominate a principle, that may serve to preside over and govern . . . several institutions or combinations of institutions that compose the matter of this science." *A Fragment on Government*, ed. Burns and Hart (Oxford: Clarendon Press, 1977), Preface, p. 416. Henry Sidgwick: "Utilitarianism may be presented as [a] scientifically complete and systematically reflective form of th[e] regulation of conduct." *Methods of Ethics*, bk. 4, ch. 3, § 1, p. 425.
97. Aristotle, *Posterior Analytics* (Cambridge, MA: Harvard University Press, Loeb Library, 1960), $71^{b}18$–23, $72^{a}15$–23, $73^{a}23$–26. There is scholarly controversy about this reading of Aristotle in light of his discussion in the *Topics* of the use of dialectic as a method of reaching first principles.
98. Aristotle, *Nicomachean Ethics*, $1095^{b}1$ff.
99. Jonsen and Toulmin, *Abuse of Casuistry*, pp. 11–19, 251–54, 296–99; Jonsen, "Casuistry as Methodology in Clinical Ethics," pp. 299–302; Brody, *Life and Death Decision Making*, pp. 12–13, 15n.

100. Toulmin, "The Tyranny of Principles," *Hastings Center Report* 11 (December 1981): 31–39.
101. Jonsen and Toulmin, *Abuse of Casuistry*, pp. 16–19.
102. Ibid., and see Toulmin, "The National Commission on Human Experimentation: Procedures and Outcomes," in *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*, ed. H. T. Engelhardt, Jr. and A. Caplan (New York: Cambridge University Press, 1987), pp. 599–613, and Jonsen, "American Moralism and the Origin of Bioethics in the United States," *Journal of Medicine and Philosophy* 16 (1991): 113–30.
103. See John D. Arras, "Getting Down to Cases: The Revival of Casuistry in Bioethics," *Journal of Medicine and Philosophy* 16 (1991): 31–33; Jonsen and Toulmin, *Abuse of Casuistry*, pp. 16–19, 66–67; Jonsen, "Casuistry and Clinical Ethics," *Theoretical Medicine* 7 (1986): 67, 71.
104. Jonsen and Toulmin, *Abuse of Casuistry*, pp. 252–62.
105. Brody, *Life and Death Decision Making*, p. 13
106. Jonsen and Toulmin sometimes seem to criticize all theory and at other times only abuse and overstatement in theory (especially as Sidgwick conceived it). We interpret them as hostile primarily to theory that is deductivist or composed of allegedly universal, eternal, and unchallengeable principles. Pragmatic and nondogmatic theories, such as those of Aristotle and William James, seem to be acceptable, even laudable. See *Abuse of Casuistry*, pp. 23–27, 279–303.
107. Jonsen, "Casuistry as Methodology in Clinical Ethics," p. 298.
108. Jonsen, "Case Analysis in Clinical Ethics," *The Journal of Clinical Ethics* 1 (1990): 65. See *Abuse of Casuistry*, p. 10.
109. Jonsen, "American Moralism and the Origin of Bioethics in the United States," esp. pp. 117, 125–28.
110. The pertinent data appear in National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, "Transcript of the Meeting Proceedings." February 11–13, 1977, pp. 11–155; July 8–9, 1977, pp. 104–17; April 14–15, 1978, pp. 155–62; and June 9–10, 1978, pp. 113–19. The Commission's general principles appear in *The Belmont Report: Ethical Guidelines for the Protection of Human Subjects* (Washington, DC: DHEW Publication (OS) 78-0012, 1978).
111. Cf. Jonsen and Toulmin, *Abuse of Casuistry*, pp. 11–16; Brody, *Life and Death Decision Making*, pp. 10–11.
112. *Utilitarianism*, ch. 2, p. 225.
113. Revisions in our theory in this edition have benefited from criticisms by David DeGrazia—see his "Moving Forward in Bioethical Theory: Theories, Cases, and Specified Principlism," *Journal of Medicine and Philosophy* 17 (October 1992): 511–39—and Ruth Faden.
114. H. A. Prichard presented powerful arguments in the common-morality tradition to show that all single or absolute-principle theories disintegrate in the face of the diversity in the considered judgments of pretheoretic commonsense morality. See his *Moral Obligation: Essays and Lectures*, ed. W. D. Ross (Oxford: Clarendon Press, 1949). However, Prichard rejected principle-based theories.
115. William K. Frankena, *Ethics*, pp. 4–9, 43–56, 113; *Thinking about Morality* (Ann Arbor: University of Michigan Press, 1980), pp. 26, 34. Frankena cites Butler on p. 6 of the first book.
116. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), p. 41.
117. Ibid., pp. 21–22.

118. Ross, *The Foundations of Ethics* (Oxford: Clarendon Press, 1939), pp. 169–70.
119. Ross, *The Right and the Good*, p. 19.
120. See Robert Nozick, "Moral Complications and Moral Structures," *Natural Law Forum* 13 (1968): 1–50.
121. Hume, "A Dialogue," published with *An Enquiry Concerning the Principles of Morals*, pp. 335–36.
122. Clouser and Gert, "A Critique of Principlism," pp. 219–27.
123. "A Critique of Principlism," p. 233.
124. Gert's moral rules can be treated as rules falling under various principles. Gert has told us in private conversation that once principles are interpreted as headings under which rules fall, they become unobjectionable, but also expendable. For the general theory on which he and Clouser rely, see Gert, *Morality: A New Justification of the Moral Rules* (New York: Oxford University Press, 1988).
125. A variant of this criticism has also been leveled against us by Ronald Green, "Method in Bioethics: A Troubled Assessment," *The Journal of Medicine and Philosophy* 15 (April 1990): 188–89.
126. See *Postures of the Mind*, pp. 139–41, 206–17, 223–26, 232–37.
127. Brandt, "Toward a Credible Form of Utilitarianism," p. 166.

3

Respect for Autonomy

Respect for the autonomous choices of other persons runs as deep in common morality as any principle, but little agreement exists about its nature and strength or about specific rights of autonomy. Many philosophers have held that morality presupposes autonomous actors, but they have emphasized different themes associated with autonomy. These disagreements indicate a need for analysis of the concept of autonomy and for specification of the principle of respect for autonomy.

The concept of autonomy is used in this chapter primarily to examine decisionmaking in health care. Our account should be adequate to identify what is protected by rules of informed consent, informed refusal, truth-telling, and confidentiality. This account is essential to our objectives throughout subsequent chapters, which fill out and qualify the nature and importance of respect for autonomy.

The Concept of Autonomy

The word *autonomy,* derived from the Greek *autos* ("self") and *nomos* ("rule," "governance," or "law"), was first used to refer to the self-rule or self-governance of independent Hellenic city-states. *Autonomy* has since been extended to individuals and has acquired meanings as diverse as self-governance, liberty rights, privacy, individual choice, freedom of the will, causing one's behavior, and being one's own person. Thus, autonomy is not a

univocal concept in either ordinary English or contemporary philosophy. Several ideas constitute the concept, creating a need to refine it in light of particular objectives. Like many philosophical concepts, "autonomy" acquires a more specific meaning in the context of a theory.

Toward this end, we start with what we take to be essential to personal autonomy, as distinguished from political self-rule: personal rule of the self that is free from both controlling interferences by others and from personal limitations that prevent meaningful choice, such as inadequate understanding.[1] The autonomous individual freely acts in accordance with a self-chosen plan, analogous to the way an independent government manages its territories and sets its policies. A person of diminished autonomy, by contrast, is in at least some respect controlled by others or incapable of deliberating or acting on the basis of his or her desires and plans. For example, institutionalized persons, such as prisoners and the mentally retarded, often have diminished autonomy. Mental incapacitation limits the autonomy of the retarded, and coercive institutionalization constrains the autonomy of prisoners.

Virtually all theories of autonomy agree that two conditions are essential: (1) *liberty* (independence from controlling influences) and (2) *agency* (capacity for intentional action). However, disagreement exists over the meaning of these two conditions and over whether some additional condition is needed. We will analyze autonomy in terms of three conditions in the next section.

Theories of Autonomy

Some theories of autonomy have featured the traits of the *autonomous person,* which include capacities of self-governance, such as understanding, reasoning, deliberating, and independent choice. However, our interest in decisionmaking leads us to focus on *autonomous choice,* which is actual governance rather than capacity for governance. Even autonomous persons with self-governing capacities sometimes fail to govern themselves in their choices because of temporary constraints imposed by illness or depression, or because of ignorance, coercion, or conditions that restrict options. An autonomous person who signs a consent form without reading or understanding the form is qualified to act autonomously by giving an informed consent, but has failed to do so. Similarly, some persons who are not generally autonomous can at times make autonomous choices. For example, some patients in mental institutions who are unable to care for themselves and have been declared legally incompetent may still be able to make autonomous choices such as stating preferences for meals, refusing some medications, and making telephone calls to acquaintances.

Some contemporary writers in ethical theory have maintained that autonomy is largely a matter of having the capacity to reflectively control and identify

with one's basic (first-order) desires or preferences through higher-level (second-order) desires or preferences.[2] For example, an alcoholic may have a desire to drink, but also a higher-order desire to stop drinking that prevails over the lower-level desire. An autonomous person, in this account, is one who has the capacity to rationally accept, identify with, or repudiate a lower-order desire or preference in a manner that it is independent of the manipulation of desires. Such acceptance of or repudiation of first-order desires at the higher level (the capacity to change one's preference structure) constitutes autonomy.

However, serious problems confront this theory of autonomy. Acceptance or repudiation of a desire on one level can be motivated by an overriding desire that is simply *stronger*, not more *rational* or *autonomous*. Second-order desires can be caused by the potency of first-order desires or by the influence of a condition such as alcohol addiction that is antithetical to autonomy. (An addicted person, we sometimes say, is not his or her own person.) The background of alcohol consumption has created the overriding desire to continue consuming. The point can be generalized beyond circumstances of addiction: If second-order desires (decisions, volitions, etc.) are generated by prior desires or commitments, then the process of identification with one desire rather than another does not distinguish autonomy from nonautonomy. Very often a second-order identification with a first-order desire is assured by the strength of the first-order desire; such identification is simply an awareness of the already formed structure of one's preferences, not a new structuring of preferences or an exercise of autonomy. Such second-order desires are not significantly different from first-order desires, and the condition of higher-order desires is an unnecessary complication for a theory of autonomy.

It also appears that any act of identification with a desire that is genuinely autonomous necessarily requires an independent act of identification at a higher level. One would have to correct, identify with, or repudiate second-order desires by third-order desires, thereby generating an infinite regress of desires, and never arriving at autonomy. If the person's identification at any level is itself the result of a process of thoroughgoing conditioning, the identification is never sufficiently independent to qualify as autonomous. For example, the alcoholic who identifies with drinking is not autonomous if a second-level desire for alcohol derives from and then reinforces the first-level desire for alcohol. An alcoholic can reflect at ever-higher levels on lower-level desires without achieving autonomy. This theory therefore at least needs an account powerful enough to distinguish autonomy-robbing influences and desires from those that are consistent with autonomy.

But this theory also needs a more important condition: It needs a way of allowing ordinary persons to qualify as deserving respect for their autonomy even when they have *not* reflected on their preferences at a higher level, and

this condition seems unlikely to be satisfied in this and various other theories of autonomy. Some theories of autonomy require that both persons and their actions meet still more rigorous standards than the standards in this theory to be autonomous. For example, some theories demand that the autonomous person be exceptionally authentic, self-possessed, consistent, independent, in command, resistant to control by authorities, and the original source of personal values, beliefs, and life plans.[3] Alternatively, some theories demand that an individual evaluate and accept each of the reasons on which the individual acts.[4] One problem with all such exacting requirements for autonomy, including second-order-desire theories, is that few choosers, and also few choices, would be autonomous if held to such standards, which in effect present an aspirational ideal of autonomy. No theory of autonomy is acceptable if it presents an ideal beyond the reach of normal choosers. Instead of depicting an ideal of this sort, our analysis will be closely tied to the assumptions of autonomy underlying moral requirements of "respect for autonomy."

We analyze autonomous action in terms of normal choosers who act (1) intentionally, (2) with understanding, and (3) without controlling influences that determine their action. The first of these three conditions of autonomy is not a matter of degree. Acts are either intentional or nonintentional. (See Chapter 4 for analysis of intention, pp. 208–210.) By contrast, the conditions of understanding and absence of controlling influences can both be satisfied to a greater or lesser extent. Actions therefore can be autonomous by degrees, as a function of satisfying these two conditions to different degrees. For both conditions, a broad continuum exists from fully present to wholly absent. Many children and many elderly patients, for example, exhibit various degrees of understanding and independence found on this continuum.[5]

For an action to be autonomous we should only require a substantial degree of understanding and freedom from constraint, not a full understanding or a complete absence of influence. To limit adequate decisionmaking by patients to the ideal of fully or completely autonomous decisionmaking strips these acts of any meaningful place in the practical world, where people's actions are rarely, if ever, fully autonomous. A person's appreciation of information and independence from controlling influences in the health care setting need not exceed, for example, a person's information and independence in making a financial investment, hiring a new employee, buying a new house, or selecting a university. Such consequential decisions are usually substantially autonomous, but far from fully autonomous.

The line between what is substantial and what is insubstantial often appears arbitrary, and therefore our analysis might seem imperiled. However, thresholds marking substantially autonomous decisions can be carefully fixed in light of specific objectives such as meaningful decisionmaking. Substantial autonomy is achievable in decisions about participation in research and acceptance

of proposed medical interventions, no less and no more than substantially autonomous choice is present elsewhere in life. Accordingly, appropriate criteria of substantial autonomy are best addressed in a particular context, rather than pinpointed through a general theory of a substantial amount.

Autonomy, Authority, and Community

Some people have argued that autonomous action is incompatible with the authority of church, state, or other communities that legislate persons' decisions. They maintain that autonomous persons must act on their own reasons and can never submit to an authority or choose to be ruled by others without losing their autonomy. Because this conclusion might seem to follow from our analysis of autonomous action, we need to consider whether autonomy is inconsistent with authority.[6]

We believe no fundamental inconsistency exists, because individuals can exercise their autonomy in choosing to accept and submit to the authoritative demands of an institution, tradition, or community that they view as a legitimate source of direction. Having welcomed the authority of his or her religious institution, a Jehovah's Witness can refuse a recommended blood transfusion, and a Roman Catholic can refuse to consider an abortion. Morality is not a set of personal rules created by individuals isolated from society, and moral principles have authority over our lives by virtue of a social and cultural setting that is independent of any single autonomous actor. That we share these principles in no way prevents them from being an individual's own principles. Virtuous conduct, role responsibilities, acceptable forms of loving, charitable behavior, respect for autonomy, and many other moral notions are autonomously accepted by individuals but usually derive from cultural traditions. A principle that is outside social arrangements would merely be one individual's belief or policy. Rules or codes of professional ethics similarly are not an individual's invention, yet they are compatible with autonomy.

This conclusion about the compatibility of autonomy, authority, and moral tradition holds for medical as well as political and religious contexts. We encounter many problems of autonomy in medical contexts because of the patient's dependent condition and the medical professional's authoritative position. On some occasions authority and autonomy are incompatible, but not because the two concepts are intrinsically incompatible. Conflict arises because authority has not been properly delegated or accepted.

Some critics of the current emphasis on autonomy in ethical theory view it as focused too narrowly on independence from others, while underestimating the importance of intimate and dependent relationships. Religious traditions are sometimes suspicious of appeals to autonomy that render the individual independent of a transcendent power, and several philosophical traditions question the model of an independent self, especially when presented as a rational will that is inattentive to communal life, reciprocity, and the development of persons over time. Some feminist critics view ethical theories that focus on autonomous individuals as unrealistic and even pernicious when a supreme and overriding value is placed on autonomy.[7] These criticisms typically apply to stark, individualistic conceptions of autonomy that more balanced theories avoid. Communal life and human relationships provide the matrix for the development of the self, and no defensible theory of autonomy denies this fact.

The Principle of Respect for Autonomy

Being autonomous is not the same as being respected as an autonomous agent. To respect an autonomous agent is, at a minimum, to acknowledge that person's right to hold views, to make choices, and to take actions based on personal values and beliefs. Such respect involves respectful *action*, not merely a respectful *attitude*. It also requires more than obligations of nonintervention in the affairs of persons, because it includes obligations to maintain capacities for autonomous choice in others while allaying fears and other conditions that destroy or disrupt their autonomous actions. Respect, on this account, involves treating persons to enable them to act autonomously, whereas disrespect for autonomy involves attitudes and actions that ignore, insult, or demean others' autonomy and thus deny a minimal equality to persons.

Why is such respect owed to persons? Two philosophers who have influenced contemporary interpretations of respect for autonomy were examined in Chapter 2: Immanuel Kant and John Stuart Mill. Kant argued that respect for autonomy flows from the recognition that all persons have unconditional worth, each having the capacity to determine his or her own destiny.[8] To violate a person's autonomy is to treat that person merely as a means, that is, in accordance with others' goals without regard to that person's own goals. Such treatment is a fundamental moral violation because autonomous persons are ends in themselves capable of determining their destinies. Mill was more concerned about the autonomy—or, as he preferred to say, the individuality—of persons in shaping their lives. He argued that citizens should be permitted to develop according to their personal convictions, as long as they do not interfere with a like expression of freedom by others; but he also insisted that we sometimes are obligated to seek to persuade others when they have false or ill-considered views.[9] Mill's position requires both noninterference with and an active strengthening of autonomous expression, whereas Kant's entails a moral imperative of respectful treatment of persons as ends rather than merely as means. In the final analysis, however, these two profoundly different philosophies both provide support for the principle of respect for autonomy.

The principle of respect for autonomy can be stated, in its negative form, as

follows: *Autonomous actions should not be subjected to controlling constraints by others*. The principle asserts a broad, abstract obligation that is free of exceptive clauses such as "We must respect individuals' views and rights *so long as their thoughts and actions do not seriously harm other persons*." Correlative to this obligation is the right to self-determination, which supports various autonomy rights, including those of confidentiality and privacy. This principle needs specification in particular contexts to become a practical guide to conduct, and appropriate specification will list the valid exceptions. Part of this process of specification will appear in rights and obligations of liberty, privacy, confidentiality, truthfulness, and consent (several of which receive sustained attention in Chapter 7). Wide disagreement exists in contemporary literature about the scope of these rights, but wide agreement exists that rights of autonomy are often legitimately constrained by the rights of others.

Respect for autonomy, then, has only prima facie standing and can be overridden by competing moral considerations. Typical examples are the following: If our choices endanger the public health, potentially harm innocent others, or require a scarce resource for which no funds are available, others can justifiably restrict our exercises of autonomy. The justification must, however, rest on some competing and overriding moral principles. The principle of respect for autonomy does not by itself determine what, on balance, a person ought to be free to know or do or what counts as a valid justification for constraining autonomy. For example, in Case 2 (see appendix) a patient with an inoperable, incurable carcinoma asks, "I don't have cancer, do I?" The physician lies, saying, "You're as good as you were ten years ago." This lie denies the patient information he may need to determine his future course of action, thereby infringing the principle of respect for autonomy. However, on balance the lie may be justified in this context by the principle of beneficence.

Many criticisms directed at current uses of the principle of respect for autonomy in biomedical ethics note that autonomy is not our only value and that respect for autonomy is not the only moral imperative.[10] These critics rightly point out that many decisions in health care depend less on respecting autonomy than on maintaining the capacity for autonomy and the conditions of meaningful life. Thus, respect for autonomy is often less prominent than expressions of beneficence and compassion. These criticisms, however, are effective only against ethical theories that recognize an unduly narrow principle of autonomy or that treat the principle as absolute or as lexically prior to all other principles. The principle of respect for autonomy should be viewed as establishing a stalwart right of authority to control one's personal destiny, but not as the only source of moral obligations and rights.

We can now consider the principle's affirmative demands, especially the *positive* obligation of respectful treatment in disclosing information and fostering autonomous decisionmaking. In some cases we are obligated to increase the

options available to persons. Many autonomous actions could not occur without the material cooperation of others in making options available. Respect for autonomy obligates professionals to disclose information, to probe for and ensure understanding and voluntariness, and to foster adequate decisionmaking. As some contemporary Kantians declare, the demand that we treat others as ends requires that we assist persons in achieving their ends and foster their capacities as agents, not merely that we avoid treating them entirely as means to our ends.[11]

There is a temptation in medicine to use the authority of the physician's role to foster or perpetuate the dependency of patients, rather than to promote their autonomy. But discharging the obligation to respect patients' autonomy requires equipping them to overcome their sense of dependence and achieve as much control as possible or as they desire. These positive obligations of respect for autonomy derive in part from the special fiduciary relationships professionals have with their patients, including affirmative obligations of disclosure and conversation. In some cases the medical professional is obligated to restore ill patients to a condition in which meaningful autonomy is possible. From this perspective, respect for autonomy is not, as some critics allege,[12] a principle that brands medical professionals as agents of paternalism and counsels them to disregard difficult patients. To the contrary, the negative and positive sides of respect for autonomy jointly indicate that respecting another includes efforts to foster and effect that person's outlook on his or her interests.

Because of various ways in which these positive and negative principles function in the moral life, they are capable of supporting many more specific moral rules (although other principles, such as beneficence and nonmaleficence, also help justify some of these same rules). Typical examples include the following rules:

1. "Tell the truth."
2. "Respect the privacy of others."
3. "Protect confidential information."
4. "Obtain consent for interventions with patients."
5. "When asked, help others make important decisions."

Both the principle of respect for autonomy and its specifications in these moral rules are prima facie, not absolute.

Despite the breadth of our obligations to respect autonomy, the principle is not so broad that it covers nonautonomous persons. The principle should not be used for persons who cannot act in a sufficiently autonomous manner (and cannot be rendered autonomous) because they are immature, incapacitated, ignorant, coerced, or exploited. Infants, irrationally suicidal individuals, and drug-dependent patients are typical examples. Those who vigorously defend rights of autonomy in biomedical ethics, as do the present authors, have never denied that some forms of intervention are justified if persons are substantially

nonautonomous and cannot be rendered autonomous for specific decisions. We will return to this problem of using respect for autonomy as a canopy to protect nonautonomous persons in this chapter's final section (and in Chapter 5 in a discussion of justifiable paternalistic interventions).

Interpreting Respect for Autonomy

The principle of respect for autonomy has been spiritedly attacked in biomedical ethics over the last several years. Many now believe that an emphasis on autonomy displaces or distorts other moral values, subverting the moral authority of medicine and leaving many patients isolated. One objective of this chapter is to interpret this complex principle so that such misconceptions and misguided criticisms are eliminated. We acknowledge, however, that many defenders of the principle have often encouraged critics because they have overextended or oversimplified it.

Varieties of "consent." The basic paradigm of autonomy in health care, politics, and other contexts is *express* and informed consent. Consent has long occupied a pivotal role in these settings because valid consent legitimates forms of authority and conduct that otherwise would not be legitimate and provides access that otherwise would be unobtainable. However, consent occurs under various conditions. It may be perfunctory or grudging, and it often occurs under intense pressures that may render it invalid. It also occurs in several forms. The paradigm of informed consent captures only one form of consent relevant to health care ethics. Another form is *tacit* consent, which is expressed passively by omissions. If residents of a long-term care facility are asked whether they object to having the time of dinner changed by one hour, a uniform lack of objection constitutes consent (assuming they understand the proposal and the need for consent). Similarly, *implicit* or *implied* consent is inferred from actions. For example, consent to one medical procedure is often implicit in a specific consent to another procedure. *Presumed* consent is still another variety, although if consent is presumed on the basis of what we know about a particular person, it closely resembles implied consent. By contrast, if it is presumed on the basis of a general theory of human goods or a theory of the rational will, the moral situation is both different and problematic. Consent should refer to an individual's own actions and inactions. Although we often legitimately make presumptions that a person's silence constitutes consent or that his or her consent is implicit in other statements or actions, these inferences can lack sufficient warrant.

Consider two situations in which non-express forms of consent—namely, implicit consent and presumed consent—are sometimes held to be morally relevant, even though their connection to autonomous choice is precarious. In current debates about organ procurement, concern exists that requirements of express consent by the decedent while alive or by the family after the relative's death impair rather than facilitate collection of needed organs. Several nations have adopted what is variously called *presumed, tacit,* or *implicit* consent for solid organs, and several states in the United States have done likewise for corneas. The moral rationale for the removal of corneas where the decedents have not registered their opposition is typically tacit consent. However, if no evidence of tacit consent exists—for example, if the decedents were unaware of the law—then the practice either falls under another form of consent or simply expropriates organs without regard to consent. It is difficult to see how consent to donation is implicit in the decedent's actions while alive, and presumed consent based on a theory of human goods is not bona fide consent. The principle of respect for autonomy, then, can be used unjustifiably through fictions of consent that are misleading and dangerous. (A related and also risky overextension is to refer to a cadaveric source of organs for transplantation as a "donor" when he or she never "donated," that is, never chose to donate.)

Another controversy about non-express forms of consent appears in the testing of hospital admittees for antibodies to the human immunodeficiency virus (HIV), which causes AIDS. Interest in testing hospitalized patients has emerged in part because treating HIV-infected patients creates risks for caregivers under some circumstances. If a newly admitted hospital patient gives express consent when asked for permission to perform an HIV-antibody test, the staff is authorized to proceed. If a patient is silent when told the test will be performed unless he or she objects, silence constitutes valid tacit consent as long as understanding and voluntariness are present. But if the question is not asked, the patient's failure to object to the test cannot be presumed to be a consent without some additional information, including the basis for presuming a general awareness of hospital testing policies.

Suppose, by contrast, that a patient has given consent for routine blood tests. Do health care professionals now have the patient's valid consent for an HIV test? Here the appeal might be to a specific consent implicit in the general consent to blood tests (or to admission to the hospital). There are reasons to be suspicious of this claim, because express consent is the only appropriate consent in these circumstances. Even if blood has already been drawn with consent, the test has psychological and social risks. For a seropositive individual, the psychological risks include anxiety and serious depression, and the social risks include stigma, discrimination, and breaches of confidentiality.[13] In general, hospitals are not justified in testing patients for HIV antibodies without specific consent.

A recent Virginia law (Virginia Code 32.1-45.1) invokes "deemed consent" in permitting a health care provider to test a patient's blood, without specific consent, following the provider's exposure to the patient's body fluids under

circumstances that could spread HIV infection: "The patient whose body fluids were involved in the exposures shall be deemed to have consented to testing for infection with human immunodeficiency virus [and] to have consented to the release of such test results to the person who was exposed." The law assigns to health care providers the responsibility to inform patients of this rule of deemed consent prior to the provision of health care services, except in emergency circumstances. The patient's acceptance of health care following this disclosure counts under the statute as consent.

Although writers in biomedical ethics often resort to fictions such as deemed consent, it is more defensible to argue straightforwardly that a patient's autonomy, liberty, privacy, or confidentiality can be justifiably overridden to obtain information about an individual's HIV-antibody status in order to protect a health care provider who has been exposed to the risk of infection. A preferable approach is to obtain patients' advance express consent if accidental exposure occurs.

Consents and refusals over time. People's beliefs, choices, and consents emerge and are modified over time. Moral and interpretive problems emerge when a person's present choices contradict his or her prior choices, which may have been explicitly designed to prevent future changes of mind. In one case a twenty-eight-year-old man decided to terminate chronic renal dialysis because of his restricted lifestyle and the burdens on his family. He had diabetes, was legally blind, and could not walk because of progressive neuropathy. His wife and physician agreed to provide medication to relieve his pain and agreed not to put him back on dialysis even if he requested this action under the influence of pain or other bodily changes as he died. While dying in the hospital, the patient awoke complaining of pain and asked to be put back on dialysis. The patient's wife and physician decided to act on the patient's earlier request not to intervene, and he died four hours later.[14] In our judgment, the spouse and physician should have put the patient back on dialysis to determine if he had autonomously revoked his prior choice.

A key question in this and related cases is whether people have autonomously revoked their prior decisions. Whether actions are autonomous can turn in part on whether they are in character or out of character. For example, a sudden and unexpected decision to discontinue dialysis by a woman who has displayed considerable courage and zest for life despite years of disability is evidence (but not necessarily decisive evidence) that a decision may not be adequately autonomous. Actions are more likely to be substantially autonomous if they are in character (for example, when a committed Jehovah's Witness refuses a blood transfusion). However, acting in character is not a necessary condition of autonomy. At most, actions that are out of character can be caution flags that warn others to seek explanations and probe more deeply into whether the actions are autonomous.

Autonomous decisions that anticipate periods of incompetence. In recent years, advance directives have emerged as a way for people to control what happens to them in the event they become incompetent. Prospective decisions to forgo life-sustaining treatment in a period of incompetence provide one example, but advance directives can have a broader role in requests for treatment and donation of organs (see pp. 241–244). One issue concerns the extent to which a person's prior autonomously expressed wishes should be viewed as valid and binding after the person becomes incompetent or dies. Suppose that subsequent to an automobile accident, a Jehovah's Witness is taken to the hospital and, while competent in the emergency room, refuses medically essential blood transfusions and falls unconscious. Some contend that the patient's prior request should be honored because otherwise the physicians (1) would have to *harm* the patient by subsequently disclosing the fact of the transfusion, or (2) would have to *deceive* the patient by not telling him the truth if they provided the transfusion.[15] On the basis of respect for autonomy, a more direct and satisfactory conclusion is available. The transfusion of an incompetent patient who, while competent, refused a transfusion violates that person's autonomy, and is disrespectful and insulting to the person.

However, situations are often far more complicated. In some cases, courts have held that in serious life-threatening emergencies only a contemporaneous informed refusal of a blood transfusion must be respected, especially if a prior directive was given in contemplation of a routine procedure that usually does not involve life-threatening complications.[16] In one case, a twenty-two-year-old man hospitalized after an automobile accident was transferred after one week to another hospital where it was determined that his severe brain injury required immediate surgery. Although his parents consented to the surgery, they refused to consent to blood transfusions. The hospital twice sought and received authority from a court to administer blood transfusions. The Pennsylvania Supreme Court held that there was no error when a lower court failed to hear testimony by the patient's parents, fiancée, or minister regarding the patient's religious beliefs, because death would probably have occurred if the operation had not been performed immediately. The necessity to preserve life outweighed third-party judgments about what an unconscious patient would want. The court held that in emergency situations calling for immediate action, "nothing less than a fully conscious contemporaneous decision by the patient will be sufficient to overrule evidence of medical necessity."[17]

Problems about personal identity and continuity. Related issues arise about whether and to what extent the prior decisions and projects of those now dead should be honored. For example, concerns exist about proper respect for the dead in autopsy, transplantation, research, and medical education, including dissection, although the dead no longer have existing interests to be protected. Similar issues surround advance directives by patients who can be expected to

survive, but never to regain competence. In Case 3 (see appendix) the staff debates whether to follow a man's request not to tell his father, who is in the early stages of Alzheimer's disease, the truth about his diagnosis. The nurse notes that the patient has a right to know and to make an advance directive prior to further deterioration, or at least to express his fears and feelings. The physician responds that the patient will lose his ability to change his mind once he loses his ability to make decisions. For the nurse, however, the patient's advance directive provides the best indication of what should later be done.

Interesting theoretical issues surrounding these controversies include personal identity and continuity of the self over time. Stated in an extreme form, the problem is that later selves can evolve to be so different from earlier selves that they are two different people. If so, it is not fair for self 1 to bind self 2 to a course of action through an advance directive—for example, when a severely demented condition causes radical changes.[18] Although this radical discontinuity thesis has some attractiveness, its plausibility and relevance in ethics are diminished when we try to envision ways to mark the point of discontinuity in order to draw the line between different selves and determine the inapplicability of advance directives. As with respect for prior wishes of the dead, we respect the previously expressed autonomous wishes of the now severely demented incompetent person because of our respect for the autonomy of the person who made the decision, as well as our own interest in securing control over our lives prior to becoming incompetent. Interventions against autonomous advance directives infringe the principle of respect for autonomy, although they can in some cases be justified.

Problems often exist in determining whether the chooser was competent when executing the advance directive, and in interpreting the directive. Similar problems about competence arise with contemporaneous choices, as we shall now see.

Competence and Autonomous Choice

Debates about competence focus on whether patients or subjects are capable, psychologically or legally, of adequate decisionmaking. Competence in decisionmaking is therefore closely connected to autonomous decisionmaking and to questions about the validity of consent.

The Gatekeeping Function

Competence judgments serve a gatekeeping role in health care by distinguishing persons whose decisions should be solicited or accepted from persons whose decisions need not or should not be solicited or accepted. Professional judgments of competence help determine whether a guardian should be appointed to look after a person's interests, whether involuntary institutionalization is appropriate, and the like. When legal incompetence has been established, a court appoints a surrogate decisionmaker who has either partial or plenary (full) authority over the incompetent individual.

In health care as well as other contexts, competence judgments distinguish the class of individuals whose autonomous decisions must be respected from those individuals whose decisions need to be checked and perhaps overridden by a surrogate.[19] No problem of competence exists if no problem exists of whether to respect a person's decisionmaking capacity. A competent decision is necessarily one for which a person can be held responsible, and a general presumption should exist that adults are competent to make decisions. Thus, the metaphor of gatekeeping is misleading if it suggests either that adults are presumed to be incompetent or that only professionals should determine whether their decisions be solicited and accepted. Moreover, if a health care professional has determined that a patient is not competent, the next step is to inquire whether capacity can be restored. When incompetence rests on a reversible cause, such as pain or overmedication, the immediate goal is to restore capacity prior to decisionmaking.

Competence judgments have a distinctive normative role of qualifying or disqualifying the person for certain decisions or actions. Even though such judgments of grading and sorting are normative, they are sometimes incorrectly presented as empirical findings. For example, a person who appears irrational or unreasonable to others might fail a psychiatric test, and so be declared incompetent. The test is an empirical measuring device, but normative judgments determine how the test will be used to sort persons into the two classes of competent and incompetent. They are normative judgments because they concern how persons ought to be or may permissibly be treated. These judgments are often justified, but they sometimes conceal an unduly narrow value perspective.

The Concept of Competence[20]

The special perspectives of medicine, psychiatry, law, philosophy, and other professions have led to competing accounts of the abilities persons must have to be competent. The word *competence* has thereby accumulated layers of meaning connected in diverse ways, but with different purposes and protective functions behind the various ideas. Consequently, some commentators believe there is no single acceptable *definition* of competence and no single acceptable *standard* of competence. They contend that no nonarbitrary line can be drawn between competent and incompetent persons. However, definitions, standards, tests, and boundary lines should be kept distinct. We confine attention for the moment to the problem of definition.

A single core meaning of the word *competence* is present in all the varied contexts in which it is properly used. That meaning is "the ability to perform a task."[21] By contrast to this core meaning, the criteria of particular competencies vary from context to context because the criteria are relative to specific tasks. The criteria for someone's competence to stand trial, to raise dachshunds, to write checks, or to lecture to medical students are radically different. The competence to decide is therefore relative to the particular decision to be made. Moreover, a person should rarely be judged incompetent with respect to every sphere of life. We usually need to consider only some type of competence, such as the competence to decide about treatment or about participation in research. These judgments of competence and incompetence affect only a limited range of decisionmaking. For example, a person who is incompetent to decide about financial affairs may be competent to decide to participate in medical research, or able to handle simple tasks easily while faltering before complex ones. Competence is therefore best understood as specific rather than global.

Many persons are incompetent to do something at one point in time and competent to perform the same task at another point in time. Judgments of competence about these persons can be complicated by the need to distinguish categories of illness that result in chronic changes of intellect, language, or memory from those characterized by rapid reversibility of these functions, as in the case of transient ischemic attack, transient global amnesia, and the like. In some of the latter cases competence varies from hour to hour.

Intermittent competence and specific competence are evident in the following case. A woman involuntarily hospitalized because of periods of confusion and loss of memory is competent most of the time to perform ordinary tasks. Health professionals and a court are called upon to determine whether this legally incompetent patient can be provided with an alternative medical therapy suitable to her situation.[22] In such legal cases, the concept of *specific incompetence* has been invoked to prevent vague generalizations about competence from excluding persons from all decisionmaking. When it proves too difficult at first to determine the level of competence, it is appropriate to evaluate the patient's understanding, deliberative capacity, and coherence over time, while supplying counseling and further support and information.

These conceptual points are of practical significance in several ways. The law has traditionally presumed that a person who is incompetent to manage his or her estate is also incompetent to vote, make medical decisions, get married, and the like. Such laws were usually aimed at the protection of property rather than persons and so were ill suited to medical decisionmaking. Their global sweep, based on a total judgment of the person, has at times been carried too far. To say that "person X is incompetent to do Y" does not necessarily imply that X is incompetent to do Z—that is, not competent to perform an action

other than Y. In one classic case, a physician argued that a patient was incompetent to make decisions because of epilepsy.[23] Such judgments defy much that we now know about the etiology of various forms of incompetence, even in hard cases of mentally retarded individuals, psychotic patients, and patients with uncontrollably painful afflictions. Persons who are incompetent by virtue of dementia, alcoholism, immaturity, and mental retardation present radically different types and problems of incompetence.

Sometimes a competent person who is generally able to select means appropriate to reach chosen goals will act incompetently in a particular circumstance. Consider the following actual case of a patient hospitalized with an acute disc problem whose goal is to control back pain. The patient has decided to manage the problem by wearing a brace, a method she has used successfully in the past. She believes strongly that she should return to this treatment modality. This approach conflicts, however, with her physician's unwavering and insistent advocacy of surgery. When the physician—an eminent surgeon who alone in her city is suited to treat her—asks her to sign the surgical permit, she is psychologically unable to refuse. The patient's hopes are vested in this assertive and, in her view, powerful and authoritative physician. Her hopes and fears are exaggerated by her illness, and she has a passive personality. In the circumstance, it is psychologically too risky for her to act as she desires. She is competent to choose in general, but she is not competent to choose on this occasion because she lacks the needed capacity.

This analysis indicates that the concept of competence in decisionmaking has close ties to the concept of autonomy, as presented earlier. A patient or subject is competent to make a decision if he or she has the capacity to understand the material information, to make a judgment about the information in light of his or her values, to intend a certain outcome, and to freely communicate his or her wish to caregivers or investigators. Law, medicine, and to some extent philosophy presume a context in which the characteristics of the competent person are also the properties possessed by the autonomous person. Although *autonomy* and *competence* are different in meaning (*autonomy* meaning self-governance; *competence*, the ability to perform a task), the criteria of the autonomous person and of the competent person are strikingly similar. Two plausible hypotheses are that an autonomous person is (necessarily) a competent person (for making decisions), and that judgments of whether a person is competent to authorize or refuse an intervention should be based on whether the person is autonomous.

Persons are more and less able to perform a task to the extent that they possess a certain level or range of abilities, just as persons are more and less intelligent and athletic. For example, an experienced and knowledgeable patient is likely to be more qualified to consent to a procedure than a frightened, inexperienced patient in the emergency room. This ability continuum runs from full

mastery through various levels of partial proficiency to complete ineptitude. However, it is confusing to think of this continuum as involving degrees of *competency*. For practical and policy reasons, we need threshold levels on this continuum below which a person with a certain level of abilities is incompetent.[24] Not all competent individuals are equally able and not all incompetent persons equally unable, but competence determinations sort persons into these two basic classes, and thus treat persons as either competent or incompetent. Where we draw the line will depend on the particular tasks involved.[25]

Although a continuum of abilities underlies the performance of tasks, the gatekeeping function of competence requires sorting persons into one of two classes: competent or incompetent. In this respect, competence is a threshold and not a continuum concept like autonomy. That is, its purpose is to divide persons into classes and not to place persons at various points on a continuum of abilities (as occurs in ratings for skating, gymnastics, and the like). Persons are not more or less competent, although they perform the tasks that determine competence more or less well. Above the threshold, persons are equally competent; below the threshold they are equally incompetent. Gatekeepers test to determine who is above and who is below this threshold.

Standards of Competence

Not surprisingly, a major question about competence in recent years has centered on standards for its determination—the conditions that must be satisfied to be judged to be competent. In law and medicine standards of competence tend to feature mental skills or capacities closely connected to the attributes of autonomous persons, such as cognitive skills and independence of judgment. In criminal law, civil law, and clinical medicine, standards for competence have clustered around various abilities to comprehend and process information and to reason about the consequences of one's actions. Although the properties most crucial to a determination of competence are controversial, in biomedical contexts a person has generally been viewed as competent if able to understand a therapy or research procedure, to deliberate regarding major risks and benefits, and to make a decision in light of this deliberation.

We agree that if a person lacks any of these capacities, competence to decide, consent, or refuse is doubtful. But there are many troublesome questions about how to classify persons who have diminished capacity to understand, deliberate, or decide. Some patients have a significant capacity to understand, deliberate, and reach conclusions without being competent. For example, some religious fanatics and many psychotic patients have fictitious and delusional beliefs that drive their actions. Nonetheless, they have considerable capacities to understand, deliberate, and decide. Patients with a low IQ as a result of meningitis at an early age can still have a significant capacity to refuse an

intervention such as intubation or catheterization. So we need further qualifications and deeper insight than our rough-and-ready categories provide.

The following case illustrates some of the difficulties often encountered in attempting to judge competence. A man who generally exhibits normal behavior patterns is involuntarily committed to a mental institution as the result of bizarre self-destructive behavior (pulling out an eye and cutting off a hand), which is influenced by his unorthodox religious beliefs. He is judged incompetent, despite his generally competent behavior and despite the fact that his peculiar actions follow "reasonably" from his religious beliefs.[26] This troublesome case cannot be interpreted in terms of intermittent competence, but it might be argued that analysis in terms of limited competence is justified. However, such an analysis suggests that persons with unorthodox or bizarre religious beliefs are less than competent, even if they reason clearly in light of their beliefs. This criterion is morally perilous for policy purposes and difficult to accept as a guideline without specific and careful qualification.

Rival standards of incompetence. The following schema expresses the range of *in*abilities currently required by various competing standards of *in*competence.[27] These standards range progressively from the one requiring the least ability to the other end of the spectrum.

1. Inability to express or communicate a preference or choice
2. Inability to understand one's situation and its consequences
3. Inability to understand relevant information
4. Inability to give a reason
5. Inability to give a rational reason (although some supporting reasons may be given)
6. Inability to give risk/benefit-related reasons (although some rational supporting reasons may be given)
7. Inability to reach a reasonable decision (as judged, for example, by a reasonable person standard)

These standards cluster around three kinds of abilities or skills (each of which requires that a level of the ability be established at which the person satisfies the standard). Standard 1 looks for the simple ability to state a preference and is a weak standard. Standards 2 and 3 probe for abilities to understand information and to appreciate one's situation. Standards 4–7 look for the ability to reason through a consequential life decision, although only standard 7 restricts the range of acceptable outcomes of a reasoning process. These standards have been and still are used, either alone or in combination, in order to determine incompetence.

Operational tests for incompetence. A clinical need also exists to select, under one or more of these standards, an operational test of incompetence that establishes passing and failing grades. Dementia rating scales, mental status exams, and similar devices test for factors such as time-and-place orientation, perseveration, memory, understanding, and coherence. These tests are empirical, clinical assessments that are generally administered when incompetence is suspected. Although empirical, a normative judgment underlies the empirical test. Each of the following ingredients involves normative judgments.[28]

1. Establishing the relevant abilities for (in)competence
2. Fixing a threshold level of the abilities in item 1
3. Accepting an empirical test for item 2

For any test accepted under item 3, it is an empirical question whether someone possesses the requisite level of abilities, but this question can only be asked and answered if other criteria have already been fixed under items 1 and 2. These criteria are sometimes fixed by institutional rules or by tradition, but in other cases they are open to further modification. Generally, even established criteria could have been different and can shift over time.

The Sliding-Scale Strategy

Properties of autonomy and psychological capacity are not the only criteria used in fashioning competence *standards*. Many policies use pragmatic criteria of efficiency, feasibility, and social acceptability to determine whether a person is competent or has given a valid authorization. For example, age has conventionally been used as an operational criterion of valid authorization, with established thresholds of age varying in accordance with a community's standards, with the degree of risk involved, and with the importance of the prospective benefits. Criteria of this type are used to protect immature or mistake-prone persons against possible decisions that fail to promote their best interest. Many such persons are competent to perform the necessary tasks and yet are judged incompetent because they fail a pragmatic standard, such as age.

In medicine, the motive for determining incompetence is to protect patients against decisions they might make that are not in their interests. Many therefore believe that standards of competence should be closely connected to levels of experience, maturity, responsibility, and welfare. Some writers offer a sliding-scale strategy for how this goal might be accomplished. They argue that as an intervention in medicine increases the risks for patients, the level of ability required for a judgment of competence to elect or refuse the intervention should be increased. As the consequences for well-being become less substantial, the level of capacity required for competence should be decreased. The sliding-scale approach allows standards of competence in decisionmaking to shift with the risk attached to the decision. If a serious risk such as death is present, then a stringent standard of competence is needed; if a low or insignificant risk is present, then a relaxed standard of competence may be used. Thus, the same person might be competent to decide whether to take a tranquillizer but incompetent to decide whether to authorize an appendectomy.[29]

The sliding-scale strategy is attractive. A decision about which standard to use to determine competence for decisionmaking depends on several factors, and these factors are often risk-related. In health care institutions, the selection of abilities, thresholds, and standards will depend on moral and policy questions involving decisionmaking requirements. All methods for setting standards of incompetence encounter difficulties about whether to emphasize the patient's autonomy or to emphasize protecting the patient against harm, a moral rather than a medical choice. If one is especially concerned about preventing abuses of autonomy, one might accept standard 1 in the previous section as the only valid standard of incompetence, or perhaps one will accept only standards 1, 2, and 3. But if one's primary concern is that sick patients receive the best medical treatment possible, one might require patients to pass all the above standards, or at least standards 6 and 7. Those who accept a stringent standard of incompetence (such as 6 and 7) will place the welfare or medical interests and safety of patients above their autonomy interests.

The strength of the sliding-scale strategy is that our interests in ensuring good outcomes legitimately contribute to the way we inquire about and create standards for judging persons competent or incompetent. If the consequences for welfare are grave, our need to be able to certify that the patient possesses the requisite capacities increases, whereas if little in the way of welfare is at stake, the level of capacity required for decisionmaking might be reduced. For example, if enteral nutrition is needed to help a patient with a reversible dementia to recover, a powerful reason exists for protecting the patient against rash or imprudent decisionmaking by adopting a rigid standard of competence. But if the dementia is irreversible and the physician's primary purpose of enteral nutrition is simply to make the patient comfortable, the standard of competence might be relaxed.

The sliding-scale strategy is generally a sound protective device, but we risk confusion about the nature of both competence judgments and competence itself unless some conceptual and moral difficulties can be resolved. This position suggests that a person's *competence* to decide is contingent upon the decision's importance or upon some harm that might follow from the decision. But this position seems questionable. A person's competence to decide whether to participate in cancer research does not depend upon the decision's consequences. As risks increase or decrease, we can legitimately escalate or reduce

the rules or measures we use to *ascertain* whether someone is competent; but in formulating what we are doing, we need to distinguish between our modes of ascertainment and the person's competence.

Two senses of *standard of competence* need to be distinguished. In one sense, *criteria* of competence are at stake—that is, the conditions under which a person is or is not competent. In a second sense, *standard of competence* refers to the *pragmatic guidelines* we use to determine competence. For example, a mature teenager could be competent to decide about a kidney transplant (satisfying criteria of competence) but could also be legally incompetent by virtue of age (failing pragmatic guidelines). In a more complicated case, a person with locked-in syndrome (involving total inabillity to communicate) is able to decide about medical care, satisfying criteria of competence, and yet fails to communicate adequately, thereby indicating through a test that the person is incompetent (failing pragmatic guidelines). To alleviate this problem of a dual meaning of *standard of competence* we will use the term *standard* only to mean a criterion for *determining* competence. Thus, a person could be correctly labeled incompetent in light of the best tests but nonetheless be competent.

Leading adherents of the sliding-scale strategy hold just the reverse—namely, that *competence* varies with risk. According to the most meticulous and convincing proponents of this strategy, Allen Buchanan and Dan Brock,

> [J]ust because a patient is competent to consent to a treatment, it does *not* follow that the patient is competent to refuse it and vice versa. For example, consent to a low-risk life-saving procedure by an otherwise healthy individual should require only a minimal level of competence, but refusal of that same procedure by such an individual should require the highest level of competence.
>
> Because the appropriate level of competence properly required for a particular decision must be adjusted to the consequences of acting on that decision, no single standard of decision-making competence is adequate. Instead, the level of competence appropriately required for decision making varies along a full range from low/minimum to high/maximal. . . . The greater the risk relative to other alternatives . . . the greater the level of communication, understanding, and reasoning skills required for competence to make that decision. . . . [30]

The core thesis in this account seems both conceptually and morally perilous. A shift in risky consequences (from risk of a small scar, say, to risk of death) indicates that we should be cautious in permitting someone to assume the greater risk, and it is true that the level of competence to decide increases as the complexity or difficulty of a task increases (deciding about spinal fusion, say, as contrasted with deciding whether to take a minor tranquilizer). But the level of competence to decide does not increase as the risk of an outcome increases. It is confusing to blend the complexity or difficulty of a task with the risk of a decision. No basis exists for believing that risky decisions require more ability at decisionmaking than less risky decisions. To the contrary, a solid basis exists for believing that many non-risky decisions require more ability at decisionmaking than many risky decisions.

Furthermore, for any person whose competence is in question, it seems disrespectful of autonomy to say, in effect, "You are competent to decide what to do with your children, what to do with your financial affairs, and whether to be in this hospital, but you are not competent to refuse to be intubated or catheterized because of the increased risk." The sliding-scale strategy seems, then, to be incoherent. Inadequate distinctions lead to a conflation of riskiness and complexity as well as conflation of criteria for justified paternalism and standards of competence.[31] These problems can be avoided by holding that the level of evidence for determining competence should vary in accordance with risk, although competence itself varies only along a scale of difficulty in decisionmaking. Brock and Buchanan insist that the "required level of decisionmaking competence" should be placed on a sliding scale from low to high in accordance with risk, but we recommend that only the required standards of evidence for determining decisionmaking competence should be placed on a sliding scale (in accordance with risk).

It follows that judgments about whether to override patients' decisions should be distinguished from questions of whether the patients are competent. Paternalism has a valid place in medicine (see Chapter 5, pp. 278–287), but its place is not in fixing criteria of competence. The incompetent are usually those we treat paternalistically, but we may also have valid paternalistic grounds for overriding the decisions of competent persons. In any event, the issue of justified paternalism should be distinguished from criteria of competence, so as to avoid situations in which we decide that a patient's decision is too risky and that he or she is therefore incompetent.

In practice, challenges to a patient's competence rarely emerge unless a disagreement exists about values. As long as the patient concurs with the physician's recommendations, his or her competence to understand, to decide, and to consent to treatment is rarely examined. But conflict between the patient's wishes and the physician's judgment about that patient's best interests typically provokes an inquiry into the patient's competence.[32] This practice is not surprising. A general presumption exists that adults are competent to make their decisions, and when those decisions are unproblematic (in part because they concur with professional judgment), no motive exists to challenge competence.

In this section we have argued that competence is determined primarily by whether a person has the capacity to decide autonomously, and not by whether a person's best interests are protected. A different set of reasons will lead us (later in this chapter and the next two chapters) to take account of risks, benefits, and best interests. First, however, we need to examine the relationship between autonomy and informed consent.

The Meaning and Justification of Informed Consent

Since the Nuremberg trials, which presented horrifying accounts of medical experimentation in concentration camps, the issue of consent has been at the forefront of biomedical ethics. The term *informed consent* did not appear until a decade after these trials, and it did not receive detailed examination until around 1972. In recent years the focus has shifted from the physician's or researcher's obligation to *disclose* information to the quality of a patient's or subject's *understanding* and *consent*. The forces behind this shift of emphasis were autonomy-driven and were primarily external to codes of medical and research ethics. Throughout this section, we note how standards of informed consent have evolved through the regulation of research, case law governing medical practice, changes in the patient–physician relationship, and ethical analysis.

Functions and Justifications of Informed Consent Requirements

Virtually all prominent medical and research codes and institutional rules of ethics now hold that physicians and investigators must obtain the informed consent of patients and subjects prior to any substantial intervention. Procedures for consent have been designed to enable autonomous choice, but they serve other functions as well, including the protection of patients and subjects from harm and the encouragement of medical professionals to act responsibly in interactions with patients and subjects.

Two positions on the function and justification of informed consent requirements have dominated the literature. Throughout the early history of concern about research subjects, consent requirements were primarily viewed as a way to minimize the potential for harm. Risk reduction and avoidance of unfairness and exploitation still function as reasons for many professional, regulatory, and institutional controls. However, in recent years the primary justification advanced for requirements of informed consent has been protection of autonomous choice, a loosely defined goal that is often buried in vague discussions of protecting the welfare and rights of patients and research subjects. Historically, we can claim little beyond the indisputable fact that a general, inchoate societal demand has developed for the protection of patients' and subjects' rights, particularly their autonomy rights. Throughout this chapter we accept and attempt to give depth to the premise that the primary function and justification of informed consent is to enable and protect individual autonomous choice.

The Definition and Elements of Informed Consent

The concept of informed consent needs clarification before we can establish the conditions under which it is appropriate and mandatory to seek informed consent. Considerable vagueness surrounds the term, creating a need to sharpen the concept so that its meaning is stable and suitable.

Some commentators have attempted to reduce the idea of informed consent to shared decisionmaking between doctor and patient, so that *informed consent* and *mutual decisionmaking* are rendered synonymous.[33] Their thesis is not that *informed consent* has this meaning in ordinary language or law, but rather that it *should* have this meaning. This proposal is plausible when consent involves ongoing exchanges of information between patients and health care providers, rather than a single event of authorizing an intervention.[34] Informed consent is typically given over time and can be withdrawn over time. We agree that it is essential to understand informed consent in terms of a temporal process, and to avoid the common view that the signed consent form is the essence of informed consent.

However, informed consent cannot be reduced to shared decisionmaking. Informed consent is obtained and will continue to be obtained in many contexts of research and emergency medicine in which shared decisionmaking is a misleading model. Furthermore, in various clinical contexts the informational exchanges through which patients elect medical interventions should be distinguished from acts of approving and authorizing those interventions. Shared decisionmaking is a worthy ideal in medicine, but it neither defines nor displaces informed consent. By viewing informed consent as a temporal process, we can avoid the model of mutual decisionmaking and the model of a single event.

Two senses of "informed consent."[35] The question "What is informed consent?" should be addressed in relation to two different conceptions of informed consent that appear in current literature and practices. In the first sense, *informed consent* is analyzable through the account of autonomous choice presented earlier in this chapter: An informed consent is an *autonomous authorization* by individuals of a medical intervention or of involvement in research. In this first sense, a person must do more than express agreement or comply with a proposal. He or she must authorize through an act of informed and voluntary consent. In the classic case of *Mohr v. Williams*, a physician obtained Anna Mohr's consent to an operation on her right ear. While operating, the surgeon determined that the left ear instead needed surgery. A court found that the physician should have obtained the patient's consent to the surgery on the left ear: "If a physician advises a patient to submit to a particular operation, and the patient weighs the dangers and risks incident to its performance, and finally consents, the patient thereby, in effect, enters into a contract authorizing the physician to operate to the extent of the consent given, but no further."[36] An *informed consent* in the first sense occurs if and only if a patient or subject, with substantial understanding and in substantial absence of control by others, intentionally authorizes a professional to do something.

In the second sense, informed consent is analyzable in terms of *the social rules of consent* in institutions that must obtain legally valid consent from patients or subjects before proceeding with therapeutic procedures or research. Informed consents are not necessarily autonomous acts under these rules and sometimes are not even meaningful authorizations. *Informed consent* refers only to an institutionally or legally effective authorization, as determined by prevailing rules. A patient or subject can *autonomously* authorize an intervention, and so give an informed consent in the first sense, without *effectively* authorizing that intervention and thus without giving an informed consent in the second sense. For example, if a minor is not legally authorized to consent, he or she can autonomously authorize an intervention without thereby giving an effective consent under prevailing institutional rules (although some "mature minor" laws give minors the right to authorize medical treatments in a limited range of circumstances).

Institutional rules of informed consent have generally not been judged by the demanding standard of autonomous authorization. As a result, critics of health care institutions' rules argue that the courts impose on physicians and hospitals nothing more than an obligation to warn of risks of proposed interventions, and that consent under these circumstances is not bona fide informed consent.[37] Jay Katz, for example, maintains that a "judicially imposed obligation must be distinguished from the idea of informed consent, namely, that patients have a decisive role to play in the medical decision making process." Katz believes that courts as well as medical institutions have allowed "the idea of informed consent . . . to wither on the vine."[38] His criticisms can best be understood in terms of the gap between the two senses of informed consent: A physician who obtains a consent under institutional criteria can fail to meet the more rigorous standards of an autonomy-based model.

It is easy to criticize institutional rules as superficial, but health care professionals cannot always obtain a consent that satisfies the demands of rigorous autonomy-protecting rules. The rules may turn out to be excessively difficult or impossible to implement. Prevailing rules should be evaluated not only in terms of respect for autonomy but also in terms of the probable consequences of imposing burdensome requirements on institutions. Policies may legitimately take account of what is fair and reasonable to require of health care professionals and researchers, the effect of alternative consent requirements on efficiency and effectiveness in the delivery of health care and the advancement of science, and the effect of consent requirements on the welfare of patients. Nevertheless, we take it as axiomatic that the model of autonomous choice (the first sense) ought to serve as the benchmark for the moral adequacy of institutional rules.

The elements of informed consent. The received approach to the definition of *informed consent* has been to specify the elements of the concept, in particular

by dividing the elements into an *information* component and a *consent* component. The information component refers to disclosure of information and comprehension of what is disclosed. The consent component refers to a voluntary decision and agreement to undergo a recommended procedure. Legal, regulatory, philosophical, medical, and psychological literatures tend to favor the following elements as the analytical components of informed consent:[39] (1) Competence, (2) Disclosure, (3) Understanding, (4) Voluntariness, and (5) Consent. These elements are then presented as the building blocks for a definition of *informed consent.* One gives an informed consent to an intervention if (and perhaps only if) one is competent to act, receives a thorough disclosure, comprehends the disclosure, acts voluntarily, and consents to the intervention.

This five-element definition is vastly superior to the one-element definition in terms of *disclosure* that courts and medical literature have often proposed,[40] but the definition is unduly influenced by medical convention and malpractice law. By making disclosure the key item, both approaches alike tend to warp informed consent. Any such definition incorporates dubious assumptions about medical authority, physician responsibility, and legal theories of liability, all of which delineate an *obligation to make disclosures* rather than a *meaning of informed consent.* The meaning of *informed consent,* as we saw above, is better analyzed in terms of autonomous authorization, which has nothing to do with disclosure specifically (even though disclosures must often be made for persons to gain an adequate understanding).

Disclosure of information is often less vital in clinical medicine than a health professional's recommendation of one or more actions. This is typically the case in direct exchanges between physicians and patients regarding surgery, medications, and the like; but it is also true, for example, of notifications to employees or pensioners by corporate medical divisions after routine surveillance or after a study of hazardous chemicals. Recommendations of treatments or of lifestyle changes such as smoking cessation are likely to be far more meaningful than information about the results of empirical studies or surveillance. Although recommendations are informational, they are also normative, and thus not appropriately described as disclosures, which are composed of descriptive statements.

Despite these reservations, we accept the premise that the above "elements" capture several basic notions about informed decision that need analysis. In this chapter each of the following seven elements is treated. (The importance of authorization in our analysis leads us to substitute elements 6 and 7 below for "Consent," which is listed above as element 5.)

I. Threshold Elements (Preconditions)

1. Competence (to understand and decide)
2. Voluntariness (in deciding)

II. Information Elements
3. Disclosure (of material information)
4. Recommendation (of a plan)
5. Understanding (of 3. and 4.)

III. Consent Elements
6. Decision (in favor of a plan)
7. Authorization (of the chosen plan)

This list requires some qualifications. First, an *informed refusal* entails a modification of items under III by turning the categories into Refusal Elements—for example, 6. "Decision (against a plan)." Second, consent for research involving human subjects does not necessarily involve a recommendation. If a recommendation is made, it may be quite different from recommendations in clinical medicine. Third, competence is more a presupposition or condition of the practice of obtaining informed consent than an element.

Because this chapter is principally about autonomy rather than informed consent and refusal, our treatment of these elements will extend into several regions of autonomous choice. We will concentrate on the elements of general importance for the analysis of both informed consent and autonomy, beginning with disclosure.

Disclosure

We have just seen that the obligation to disclose information to patients has often been presented as a necessary, and sometimes as the sole, condition of valid informed consent. The legal doctrine of informed consent has been primarily a law of disclosure based on a general obligation to exercise reasonable care by giving information. Civil litigation has emerged over informed consent because of injury to one's person or property that is intentionally or negligently inflicted by a physician's failure to disclose, an injury measured in terms of monetary damages. This focus results from the legal system's need for a serviceable mechanism to assess injury and responsibility.

As litigation over legal requirements of consent to medical treatment evolved, a more complicated set of rules also evolved, especially regarding disclosure standards. The term *informed consent* was born in this legal context. However, from the moral viewpoint, informed consent has less to do with the liability of professionals as agents of disclosure and more to do with the autonomous choices of patients and subjects. Both health care professionals and patients need to ask and answer questions, and this process is less a matter of disclosing information than of discovering the relevant information and deciding how to frame and use it.

Nevertheless, disclosure is a pivotal topic. Without an adequate way to deliver information, many patients and subjects will have a substandard basis for decisionmaking. The professional's perspective, opinions, and recommendations are often essential for a sound decision. Professionals are obligated to disclose a core set of information, including (1) those facts or descriptions that patients or subjects usually consider material in deciding whether to refuse or consent to the proposed intervention or research, (2) information the professional believes to be material, (3) the professional's recommendation, (4) the purpose of seeking consent, and (5) the nature and limits of consent as an act of authorization.

Additional types of disclosure have also been proposed—for example, the names of persons in charge. Many controversies center on how much should be disclosed about a procedure's risks, as well as about its nature and its benefits, and any alternatives to the procedure, including new drugs, devices, and treatments.[41] If research is involved, disclosures should generally be made as to the aims, methods, anticipated benefits and risks of the research, any anticipated inconvenience or discomfort, and the subjects' right to withdraw from the research. Additional disclosures and special precautions to ensure that persons understand may also be necessary, including disclosure of the criteria used for the selection of subjects and an indication that the person has an opportunity to ask further questions. Such lists could be expanded almost indefinitely. For example, in one controversial decision, the California Supreme Court held that when seeking an informed consent "a physician must disclose personal interests unrelated to the patient's health, whether research or economic, that may affect the physician's professional judgment."[42] However, the central moral issues turn on the informational needs of particular patients and subjects, not on lists or categories of information.

Standards of Disclosure

The courts have struggled to determine which norms should govern the disclosure of information. Two competing standards of disclosure have emerged: the professional practice standard and the reasonable person standard. A third, the subjective standard, has also been proposed, but it has been implemented in the courts only as a causation standard—that is, as a way of determining whether a physician's failure to disclose caused injury to the patient—not as a disclosure standard. Discussion of these standards has set the terms for much of the current debate over informed consent requirements in medical ethics no less than in law.

The professional practice standard. The first standard holds that adequate disclosure is determined by a professional community's customary practices. This standard assumes that the doctor's proper role is to act in the patient's best

medical interest. The custom in a profession establishes the amount and kinds of information to be disclosed. Disclosure, like treatment, is a task that belongs to physicians because of their professional expertise and commitment to the patient's welfare. As a result, only expert testimony from members of this profession could count as evidence that there has been a violation of a patient's right to information.[43]

Several difficulties affect this standard, which is sometimes called a *reasonable doctor standard*.[44] First, it is uncertain in many situations whether a customary standard exists for the communication of information in medicine. Second, if custom alone were conclusive, pervasive negligence could be perpetuated with impunity. The majority of professionals could offer the same inadequate level of information or be allowed total discretion to determine the scope of disclosure. The chief objection to the professional practice standard is that it subverts the right of autonomous choice. Professional standards in medicine are fashioned for medical judgments, but decisions for or against medical care, which are nonmedical, are rightly the province of the patient.

It is also questionable whether physicians have developed skills to determine the information that is in their patients' best interests. The assumption that they have such expertise is largely empirical, and yet no reliable data substantiate it.[45] The weighing of risks in the context of a person's subjective beliefs, fears, and hopes is not an expert skill, and information provided to patients and subjects sometimes needs to be freed from the entrenched values and goals of medical professionals.

The reasonable person standard. Although many legal jurisdictions retain the traditional professional practice standard, the reasonable person standard has gained some acceptance in over half of the states in the United States.[46] According to this standard, information to be disclosed is determined by reference to a hypothetical reasonable person. The pertinence of information is measured by the significance a reasonable person would attach to it in deciding whether to undergo a procedure. The authoritative determination of informational needs thus shifts from the physician to the patient, and physicians may be found guilty of negligent disclosures even if their behavior conforms to recognized professional practice. Proponents of the reasonable person standard believe that obligations to respect autonomy generally outweigh obligations of beneficence and that, on balance, the reasonable person standard better serves the autonomy of patients than does the professional practice standard.

Whatever its merits, this reasonable person standard is also plagued by conceptual, moral, and practical difficulties. First, "material information" and the central concept of the reasonable person have never been carefully defined. Second, questions exist about whether and how the reasonable person standard

can be employed in practice. Its abstract and hypothetical character makes it difficult for physicians to use, because they have to project what a reasonable patient would need to know. A related problem has emerged from empirical studies examining whether patients use information disclosed to them in reaching their decisions. Data collected in one study indicate that although ninety-three percent of the patients surveyed believed they benefited from the information disclosed, only twelve percent used the information in their decisions to consent.[47] This study, involving family-planning patients, reaches conclusions similar to an earlier study of kidney donors.[48] In both studies the data indicate that patients generally make their decisions prior to and independent of the process of receiving information. Other studies indicate that patients often deferentially accept physicians' or parents' recommendations without carefully weighing risks and benefits,[49] and that many patients would agree to a procedure without any discussion of risks (eighty-six percent of upper gastrointestinal endoscopy patients in one study)[50] or during their first meeting with a physician (eighty-two percent of candidates for breast cancer adjuvant therapy in one study).[51]

These data do not always indicate that decisions by patients are uninformed or that disclosed information is irrelevant. The patients may have believed that additional information received from physicians did not alter their prior commitment to a course of action such as surgery. Nonetheless, these empirical findings raise questions about what should count as material information for the individual patient and whether this information is the same for the reasonable patient and the individual patient. This problem leads to the third standard of disclosure.

The subjective standard. In the subjective model, adequacy of information is judged by reference to the specific informational needs of the individual person, rather than the hypothetical "reasonable person." Individual needs can differ, because persons may have unconventional beliefs, unusual health problems, or unique family histories that require a different informational base than the reasonable person needs. For example, a person with a family history of reproductive problems might desire information that other persons would not want or need before becoming involved in research on sexual and familial relations or accepting employment in certain industries. If a physician knows or has reason to believe that a person wants such information, then withholding it may undermine autonomy.

At issue is the extent to which a standard should be tailored to the individual patient, that is, made subjective, so that a disclosure would have to include the factors particular to the patient's need for information that a physician could reasonably be expected to know. According to the subjective standard, the phy-

sician is obligated to disclose information a particular patient needs to know, so long as a reasonable connection exists between those needs and what the physician should know about the patient's position.[52]

Despite many problems that plague the subjective standard as a *legal* standard, it is a preferable *moral* standard of disclosure, because it alone acknowledges the independent informational needs of persons. Nevertheless, exclusive use of a subjective standard is insufficient for either law or ethics, because patients often do not know what information would be relevant for their deliberations, and a doctor cannot reasonably be expected to do an exhaustive background and character analysis of each patient to determine what information would be relevant. Again, the key question is not what quantum of information should be disclosed, but what professionals can do to facilitate informed decisionmaking.

The solution to the problem of disclosure should be found in active participation through mutual exchange of information. Neither the professional-practice standard nor the reasonable-person standard is a sufficient guide. What professionals customarily disclose and what an objective reasonable person needs often fail to contain some or all of the information material to the person making the decision. Legal and professional rules of disclosure should only serve, then, to initiate the communication process, and professionals and their institutions should not be satisfied with a signed consent form unless attention has also been paid to the process that led to it.

Intentional Nondisclosure

Several problems in biomedical ethics center on deliberate nondisclosure. Some types of research are incompatible with complete disclosure, and in certain clinical interventions physicians often claim that nondisclosures benefit the patient. Are such intentional nondisclosures justifiable?

The therapeutic privilege. Legal exceptions to the rule of informed consent allow the health professional to proceed without consent in cases of emergency, incompetency, waiver, and the like. A controversial exception is the therapeutic privilege, according to which a physician may legitimately withhold information, based on a sound medical judgment that to divulge the information would be potentially harmful to a depressed, emotionally drained, or unstable patient. Several harmful outcomes have been cited, including endangering life, causing irrational decisions, and producing anxiety or stress.[53] Despite the protected status this doctrine has traditionally enjoyed, in 1986 U. S. Supreme Court Justice Byron White vigorously attacked the idea that concerns about increasing a person's anxiety about a procedure provide grounds for an exception to rules of informed consent: "It is the very nature of informed consent provisions that they may produce some anxiety in the patient and influence her in her choice. This is in fact their reason for existence, and . . . it is an entirely salutary reason."[54] White suggested that the legal status of the doctrine of therapeutic privilege is no longer as secure as it once was.

The precise formulation of this therapeutic privilege varies across legal jurisdictions. Some formulations permit physicians to withhold information if disclosure would cause *any* countertherapeutic deterioration in the patient's condition. Other formulations permit the physician to withhold information if and only if the patient's knowledge of the information would have serious health-related consequences—for example, by jeopardizing the treatment's success or by critically impairing relevant decisionmaking processes. The narrowest formulation is analogous to a circumstance of incompetence: The therapeutic privilege can be validly invoked only if the physician has reason to believe that disclosure would render the patient incompetent to consent to or refuse the treatment. To invoke the therapeutic privilege under this condition does not in principle conflict with respect for autonomy, because the patient would not be capable of an autonomous decision at the point it would be needed.

It is doubtful, however, that a general criterion of physician discretion in disclosing information can be made coherent with the extensive disclosure of facts now demanded by many courts. Courts rarely mention valid exceptions to the legal rights they proclaim, and they also leave unclear whether remote risks—often the most anxiety-inducing—must be discussed. The moral issues are as tangled as the legal issues, as we can see in a case in which a woman had a fatal reaction during urography. The radiologist had intentionally not disclosed the chance of death (roughly one in ten thousand) because it might have upset the patient. The radiologist justified his nondisclosure on grounds that the disclosure would be "dangerous" and "not in the best interest of the patient."[55] Such nondisclosure may be warranted with exceptionally fragile patients, but the cases will be rare. Empirical evidence indicates more often than not that physician-hypothesized negative effects such as anxiety and reduced compliance do not materialize.[56] A more acceptable approach is found in the following recommendation by a group of anesthesiologists: "Tell all patients that there are serious, although remote, risks of anesthesia, but . . . allow the individual patient to decide how much additional information he or she wishes to obtain about these risks."[57]

Therapeutic use of placebos. The therapeutic use of placebos also involves intentional deception or incomplete disclosure. A placebo is a substance or intervention that the health care professional believes to be pharmacologically or biomedically inert for the condition being treated. Studies indicate that pla-

cebos relieve some symptoms in approximately thirty-five percent of patients who suffer from conditions such as angina pectoris, cough, anxiety, depression, hypertension, headache, and the common cold.[58]

One incautious, beneficence-based defense of placebos is that "deception is completely moral when it is used for the welfare of the patient."[59] This defense endangers autonomy and may founder on its assumptions. Some evidence suggests that the placebo effect—an improvement in the patient after use of a placebo—can sometimes be produced without nondisclosure, incomplete disclosure, or deception. For example, the placebo effect sometimes occurs even if patients have been informed that a substance is pharmacologically inert and have consented to its use.[60] In many cases placebos appear to work because of the "healing context," involving the professional's care, compassion, and skill in fostering hope and trust.[61] Thus, the placebo effect is sometimes produced without administering placebos.

Nevertheless, a placebo is less likely to be effective if used with the patient's knowledge. In one case, professionals thought that an undisclosed placebo offered the only hope of effective pain treatment. Mr. X had undergone several abdominal operations for gallstones, postoperative adhesions, and bowel obstructions, and subsequently experienced chronic pain. He became somewhat depressed, lost weight, had poor personal hygiene, was unkempt, and withdrew socially. After using the addictive drug Talwin six times a day for more than two years to control pain, he had trouble finding injection sites for the Talwin. He sought help "to get more out of life in spite of pain" and voluntarily entered a psychiatric ward that used relaxation techniques and other behavioral procedures.

In the ward he successfully reduced his Talwin usage to four times daily, but insisted that this level was necessary to control his pain. His therapists decided to withdraw the Talwin over time without his knowledge by diluting it with increasing proportions of normal saline. He experienced withdrawal symptoms of nausea, diarrhea, and cramps, which he thought were the result of the Elavil, which the therapists had introduced to relieve the withdrawal symptoms, again without informing him of the purpose. The physicians gradually increased the intervals between injections of the saline. X was aware of these changes, but did not know that the injections contained only saline. After three weeks, his therapist informed him of the placebo substitution. After his initial incredulity and anger subsided, the patient asked that the saline be discontinued and self-control techniques continued. When discharged three weeks later, he could control his abdominal pain more effectively with the self-control techniques than he had been able to with Talwin. Six months later he was still using self-control techniques and had resumed social activities.

The therapists defended their deception on the grounds that they "felt ethically obliged to use a treatment that had a high probability of success." Yet,

the therapists continue, "We saw no option without ethical problems. Although it is precarious to justify the means by the end, we felt most obliged to use a procedure designed to help the patient achieve a personally and medically desirable goal." They argued, in effect, that the principle of beneficence overrode respect for autonomy in this case. They also hinted that their actions did not infringe X's autonomy because his autonomy was already compromised by his addiction. They further suggested that they did not violate X's rights, and even acted in accord with his autonomous choices by taking account of his own therapeutic goals.[62]

One defense of this claim appeals to X's alleged implicit consent when he admitted himself to a ward where adjustment in medication was a clear expectation. He accepted the therapy, "to get more out of life." However, what X implicitly consented to when he entered the psychiatric ward is unclear, because we do not know what he understood, and no appeal to implicit consent can be accepted for all maneuvers, because he expressly refused to allow further reduction in his Talwin dosage. A related justification for the undisclosed placebo is that X ratified the therapists' decision to use the placebo when he decided to continue the self-control techniques, rather than returning to Talwin. However, predicted future "consent" is only an expected future approval, not a *consent,* and at best rests on evidence about the patient's beliefs or goals. But even with such evidence, a predicted future ratification does not transform the current intervention into respect for the patient's autonomy.

The therapists note that they saw "no option without ethical problems," but they had not exhausted all moral options. One possibility was to obtain the patient's general consent to the administration of several drugs and placebos, as part of the effort to wean him from Talwin and to enable him to develop adequate self-control techniques to manage his pain. Such consent, obtained at the outset, would have obviated the need for specific consent to the placebo substitution. The staff's commitment to behavioral therapy may have blinded them to some aspects of the problem, leading them to focus on correctable behaviors, with less attention to the person behind the behaviors.[63] If the therapists had conceived X's problem in terms of the importance of autonomous choice, they might have discovered alternative procedures with fewer or no ethical problems. We conclude that the staff's justification is deficient in its appeals to implicit consent and future consent. However, we leave open the possibility that a paternalistic use of placebos of the sort found in this case may be justifiable on other grounds (see Chapter 5).

Withholding Information from Research Subjects. Problems of intentional nondisclosure in clinical practice have parallels in research, where investigators sometimes need to avoid sharing available information with subjects. Occasionally good arguments can be provided for such nondisclosure. Vital research in

fields such as epidemiology could not be conducted if consent from subjects were required to obtain access to medical records, and their use without consent is sometimes ethically justified, for example, to establish the prevalence of a particular disease. Some research of this description is only the first phase of an investigation intended to determine whether a need exists to trace and contact particular individuals who are at risk of disease and obtain their permission for further participation in a study. There should be careful protection of confidential information used, but informed consent requirements are sometimes overly burdensome. Occasionally research subjects need not be contacted at all, for example, when hospital records are studied by epidemiologists without knowing the names of the patients. And in other circumstances persons need only to be *notified* in advance of how data being gathered will be used and given the opportunity to refuse participation in the research. That is, disclosures, warnings, and opportunities to decline involvement are sometimes legitimately substituted for obtaining an informed consent.

However, data collection and its analysis grow in unanticipated ways over time, and results occasionally appear late in the course of research that could not be anticipated at the beginning. A conscientious investigator will periodically consider whether informed consent is needed, and submit these judgments to a review committee. Such review may discover a need for informed consent where none previously existed or find a need to communicate study results to those whose records have been examined.

Many forms of intentional nondisclosure in research are difficult to justify. For instance, debate has emerged about a recent study, designed and conducted by two physicians at the Emory University School of Medicine to determine the prevalence of cocaine use and the reliability of self-reports by patients. Controversy centered on the questions that would be most likely to elicit accurate answers from a group of men in an Atlanta walk-in, inner-city hospital clinic that serves low-income, predominantly black residents. In this study, which was approved by the institutional human investigations committee, weekday outpatients at Grady Memorial Hospital were "asked to participate in a study about asymptomatic carriage of STDs," for which they would receive ten dollars. Study participants had to be between 18 and 39 years old and had to be sexually active within the previous six months. Eighty-two percent of those asked agreed to participate. The average age of the participants was 29.5 years; 91.6% were black, and 89% were uninsured.

The participants provided informed consent for the sexually transmitted disease study, but not for the unmentioned piggy-back study of recent cocaine use and reliability of self-reports of such use. Patients were informed that their urine would be tested for STDs, but they were not informed that it would be analyzed for cocaine metabolites. Of the 415 eligible men who agreed to participate, thirty-nine percent tested positive for a major cocaine metabolite, although seventy-two percent of those with positive urinary assays denied any illicit drug use in the three days prior to sampling. (The metabolite is cleared from the body within three days following single-dose cocaine injection.) In answering questions, subjects with positive urine assays were more likely to admit to "any illegal drug" use (87.5 percent) than admit to the more specific "any form of cocaine" use (60.6 percent) over the prior year. Overall, "42.4 percent of the 415 participants admitted to use of cocaine within one year." Researchers concluded:

> Our findings underscore the magnitude of the cocaine abuse problem for young men seeking care in inner-city, walk-in clinics. Health care providers need to be aware of the unreliability of patient self-reports of illicit drug use. In this high-risk population, any admission of illicit drug use within the prior year, despite denial of ongoing abuse, should lead physicians to suspect recent use.[64]

The researchers deceived the subjects about some aims and purposes of the research (to study the prevalence of recent cocaine use and the reliability of patient self-reports) and did not disclose the means that would be used (testing their urine for recent cocaine use). Investigators faced a dilemma. On the one hand, accurate information was needed about illicit drug use for health care and public policy. On the other hand, obtaining adequate informed consent was difficult. If informed about the aims of the study, many potential subjects would either refuse to participate or would offer false information to researchers. These problems are increased in research that uses subjects' body fluids as a way to test the reliability of subjects' self-reports about illegal activities.

Requirements of informed consent should not be so easily set aside. These rules protect subjects from manipulation and abuse during the research process. For example, reports of this cocaine study stand to increase suspicion of medical institutions and professionals and could also function to make patients' self-reports of illegal activities even less reliable.[65] We have conceded that informed consent is sometimes unnecessary, but this study of cocaine use is not a legitimate example. Investigators would have been better advised to resolve their dilemma by developing alternative research designs, including sophisticated methods of using questions that can either reduce or eliminate response errors without abridging informed consent. (Another example of problematic partial nondisclosure appears in some clinical trials involving randomization, placebos, blind experiments, and the like; see Chapter 7, pp. 442–447).

As substantial deception or substantial risk is added in a research project, justifying the research becomes progressively more difficult. In Stanley Milgram's well-known experiments on obedience, recruits come to what appears to be a psychology laboratory to participate in what has been advertised as a study of memory and learning. One is designated a "teacher," the other a "learner." The investigator explains that the study will focus on how punish-

ment affects learning. The learner is then seated, his or her arms are strapped to prevent excessive movement, and an electrode is attached to his or her wrist. The learner is told that he or she is to attempt to learn a list of word pairs. Whenever an error is made, electric shocks of increasing intensity will be administered by the other subject, the teacher. The subject assigned the role of teacher is deliberately deceived by the investigator. The learner is actually part of the research team, and the machine does not deliver shocks to the learner. The point of the experiment, as Milgram describes it, "is to see how far a person will proceed in a concrete and measurable situation in which he is ordered to inflict increasing pain on a protesting subject."[66]

In contrast to what consultants had predicted, Milgram reported that as many as 62.5 percent of the subjects continued to obey the investigator's order to inflict shock up to the maximum of 450 volts, labeled on the machine "Danger—Severe Shock." Through debriefing and friendly reconciliation with the "victim," Milgram tried to minimize any harms that the subjects might experience through stress, anxiety, guilt, and shame about their actions in the experiment. His primary defense of his methods is that the subsequent responses of the participants justify the research: "To my mind, the central moral justification for allowing my experiment is that it was judged acceptable by those who took part in it."[67] Eighty-four percent said they were glad to have been in it, fifteen percent were neutral, and one percent expressed negative opinions. These subsequent responses could be construed as indicating that the harms were minimal (and justified by the benefits) and as providing a form of retroactive approval and consent.

The extensive ethical debate over this research during the last thirty years has centered on the imposition of risks of harm to subjects without their informed consent.[68] Milgram contends that criticisms of his methods are unfair if the research is described as involving "deception." He proposes "morally neutral terms" such as "masking," "staging," and "technical illusions." However, these terms frustrate the debate by obscuring the fact that deception without the subjects' specific consent was essential for the research. The experiment may have eventuated in subsequent approval by subjects in their efforts to come to grips with their actions, but retroactive approval is not a substitute for informed consent or refusal. A more promising approach, as Milgram recognizes, is obtaining subjects' prior general consent to participate in research involving deception or nondisclosure.[69]

We believe that research cannot be justified if (1) significant risk is involved, and (2) subjects are not informed that they are being placed at risk. Again, the central question is whether subjects can voluntarily accept the risk with an adequate understanding of the deceptive practices. It is, in our judgment, an indefensible violation to deceive subjects while placing them at substantial risk, regardless of the research's importance. This conclusion does not imply that

research involving deception can *never* justifiably be undertaken. Relatively risk-free research that requires deception or incomplete disclosure is often warranted in fields such as behavioral and physiological psychology, as well as the biomedical sciences. Simple examples include studies of visual and other perceptual responses, as well as some behavioral observation studies. Generally, however, deception should be permitted in research only if it is essential to obtain vital information, no substantial risk is involved, subjects are informed that deception is part of the study, and subjects consent to participate under these conditions. We return in Chapter 7 to some needed qualifications to this conclusion, especially in the context of randomized clinical trials.

Understanding

Traditional problems of disclosure need to be reconceived in terms of what professionals can do to facilitate good decisions based on substantial understanding. Asking questions, eliciting the concerns and interests of the patient or subject, and establishing a climate that encourages questions often do more to foster understanding than disclosed information does. However, clinical experience and empirical data indicate that patients and subjects exhibit wide variation in their understanding of information about diagnoses, procedures, risks, and prognoses. Some patients and subjects are calm, attentive, and eager for dialogue, whereas others are nervous or distracted in ways that impair or block understanding. Many conditions limit their understanding, including illness, irrationality, and immaturity.

The Nature of Understanding

No consensus exists about the nature of understanding, but an analysis sufficient for our purposes is that one understands if one has acquired pertinent information and justified, relevant beliefs about the nature and consequences of one's action. Such understanding need not be complete, because a substantial grasp of central facts and other descriptions is generally sufficient. Some facts are irrelevant or trivial; others are vital, perhaps decisive. In some cases, a person's lack of awareness of even a single risk, limitation, or missing fact can deprive him or her of adequate understanding. Consider, for example, the case of *Bang v. Miller Hospital,* in which patient Bang did not intend to consent to a sterilization entailed in prostate surgery.[70] Bang did, in fact, give a formal consent to prostate surgery, but without being told that sterilization was an inevitable outcome. (Sterilization is not necessarily an outcome of all prostate surgery, but it is inevitable in the specific procedure selected in this case.) Bang's failure to understand this one surgical consequence substantially com-

promised what was otherwise an adequate understanding and invalidated what otherwise would have been a valid consent.

Patients and subjects usually should understand at least what a health care professional believes a patient or subject needs to understand and should regard as material in order to authorize an intervention. Typically, diagnoses, prognoses, the nature and purpose of the intervention, alternatives, risks and benefits, and recommendations are essential. But patients or subjects also need to share an understanding with professionals about the terms of the authorization before proceeding. As in all contractual circumstances, unless agreement exists about the essential features of what is authorized, there can be no assurance that a patient or subject has made an autonomous decision. Even if both physician and patient use a word such as *stroke* or *hernia*, their interpretations will be vastly different if standard medical definitions and conceptions have no meaning or significance for the patient.

It is sometimes argued that many patients and subjects cannot comprehend enough information or appreciate its relevance sufficiently to make decisions about medical care or participation in research. Franz Ingelfinger maintains, for example, that "the chances are remote that the subject really understands what he has consented to."[71] Such statements are overgeneralizations based partially on unwarranted standards of full disclosure and full understanding. The ideal of complete disclosure of all possibly relevant knowledge promotes these claims about the limited capacity of subjects to comprehend. If this ideal standard is replaced by a more acceptable account of understanding relevant information, such skepticism can be put to rest. From the fact that actions are never *fully* informed, voluntary, or autonomous, it does not follow that they are never *adequately* informed, voluntary, or autonomous.

Some patients, however, have such limited knowledge bases that communication about alien or novel situations is exceedingly difficult, especially if new concepts and cognitive constructs are required. Studies indicate that their understanding of scientific goals and procedures is likely to be both impoverished and distorted.[72] But even under such difficult situations, enhanced understanding and adequate decisions are often possible. Successful communication of novel and specialized information to laypersons can often be accomplished by drawing analogies between this information and more ordinary events familiar to the patient or subject. Similarly, professionals can express risks in both numeric and nonnumeric probabilities, while helping the patient or subject to assign meanings to the probabilities through comparison with more familiar risks and prior experiences, such as risks involved in driving automobiles or using power tools.

However, to enable a patient not only to comprehend but also to appreciate risks and benefits can be a formidable task. For example, many patients confronted with coronary artery bypass, orthopedic operations, and many other

forms of surgery understand that as a consequence of their consent to the surgery they will suffer postoperative pain. Nevertheless, their projected expectations of the pain are often altogether inadequate. Patients often cannot, in advance, adequately appreciate the nature of the pain, and many ill patients reach a point at which they can no longer balance with clear judgment the threat of pain against the risks of surgery. At this point the benefits of surgery are overwhelmingly attractive, and the risks are devalued. In one respect these patients correctly understand basic facts about procedures that involve pain, but in other respects their understanding is less than adequate.

Many situations in medicine require physicians to confront this problem. For example, in Case 4 (see appendix), a fourteen-year-old girl consents to donate a kidney to her mother. Although she has exhibited a perceptive and relatively unemotional grasp of the situation, many doubt that a fourteen-year-old child can in these circumstances either adequately appreciate the significance of future risks or carefully balance risks and benefits.

Problems of Information Processing

With the exception of a few limited studies of comprehension, studies of decisionmaking by patients typically pay little attention to information processing, which raises substantial issues about understanding. For example, information overload is sometimes an obstacle to adequate understanding and is as likely as underdisclosure to produce uninformed decisions. Information overload is exacerbated if unfamiliar terms are used or if information cannot be meaningfully organized, yet practical constraints generally require that disclosures occur in a compact presentation. Patients and potential subjects are likely to rely on some modes of selective perception, and it is often difficult to determine when words have special meaning for them, when preconceptions distort their processing of the information, and when other biases intrude.

Some valuable studies have uncovered difficulties in processing information about risks, indicating that risk disclosures often lead subjects to distort information and promote inferential errors and disproportionate fears of risks.[73] Some ways of framing information are so misleading that both health professionals and their patients regularly distort the content. For example, choices between risky alternatives can be heavily influenced by whether the same risk information is presented as providing a gain or an opportunity for a patient, or as constituting a loss or a reduction of opportunity.[74] One study asked radiologists, outpatients with chronic medical problems, and graduate business students to make a hypothetical choice between two alternative therapies for lung cancer: surgery and radiation therapy.[75] The preferences of all three groups were affected by whether the information about outcomes was framed in terms of survival or death. When faced with outcomes framed in terms of probability

of *survival*, twenty-five percent chose radiation over surgery. However, when the identical outcomes were presented in terms of probability of *death*, forty-two percent preferred radiation. The mode of presenting the risk of immediate death from surgical complications, which has no counterpart in radiation therapy, appears to have made the decisive difference.

These framing effects reduce understanding of material information, with direct implications for autonomous choice. If a misperception prevents a person from adequately understanding the risk of death and this risk is material to the person's decision, then the person's choice of surgery is based on less than a substantial understanding and would not qualify as an autonomous authorization. The lesson to be learned is not skepticism about information processing, but rather the need for better understanding of techniques that will enable professionals to communicate the positive and the negative sides of information—for example, both the mortality and the survival information.[76]

Problems of Nonacceptance and False Belief

A person's ability to make decisions can be compromised by a breakdown in the ability to *accept* information as true or untainted, even if the person adequately *comprehends* the information. The distinction between comprehension of information and acceptance of information has often been obscured in the relevant literature by excessive reliance on recall tests. At best, "correct" answers on these tests provide evidence of a person's memory of what the physician or investigator has disclosed, not whether the subject interpreted correctly or believed what was disclosed.

However, a false belief can invalidate a decision by a patient or subject in the presence of suitable disclosure and comprehension. Here are three examples: (1) A person might falsely and irrationally believe that a doctor will not fill out insurance forms unless the patient consents to a procedure the doctor has suggested. (2) A sufficiently informed psychiatric patient capable of consent might agree to participate in nontherapeutic research under the false belief that it is therapeutic. (3) A seriously ill patient asked to make a treatment decision might refuse under the false belief that he or she is not ill. Even if the physician recognizes the person's false belief and adduces conclusive evidence to prove to the person that the belief is mistaken, and the person comprehends the information provided, the person may go on believing that what has been (truthfully) reported is false.

Inconclusive evidence and failure to achieve agreement about the truth or falsity of beliefs, sometimes after considerable discussion, further complicate these problems. Many beliefs that are central for a patient's decision are regarded by others, including health care professionals seeking consent, as highly questionable, poorly reasoned, or perhaps absurd. Sometimes overwhelming

medical evidence indicates that a patient's belief is unjustified, but in other circumstances the patient's beliefs will be contestable without being refutable by hard counterevidence.

The probabilities and uncertainties that surround many beliefs suggest that truth claims should be judged by the available evidence, which is often subject to different interpretations. More than one standard of evidence may exist, and all evidence must be collected within some framework that determines what qualifies as evidence. No evidence is independent of the framework that it presupposes, yet two or more frameworks sometimes advance competing standards of evidence. If disagreement persists on the criteria for determining the justifiability of beliefs, there will be no adequate basis for determining whether a given belief compromises understanding or simply involves an essentially contestable proposition. This conclusion is not meant as a skeptical denial of the possibility of knowledge, but only as a warning that the evidence for thinking that a belief is false may be rationally contestable.

When beliefs are demonstrably false, the question arises whether patients and subjects should be forced to abandon their false beliefs to enable them to reach an informed decision. Some have argued that if autonomous subjects decline further information, it should not be imposed.[77] This proposal is attractive, but it also seems wrong to say that we should never pressure protesting patients or subjects to change their beliefs or to process information differently. If choice is limited by ignorance, as in the case of a demonstrably false belief, it may be permissible or possibly obligatory to promote autonomy by attempting to impose unwelcome information.

Consider the following case in which a false belief played a major role in a patient's refusal of treatment:[78]

> A 57-year-old woman was admitted to the hospital because of a fractured hip. . . . During the course of the hospitalization, a Papanicolaou test and biopsy revealed stage 1A carcinoma of the cervix. . . . Surgery was strongly recommended, since the cancer was almost certainly curable by a hysterectomy. . . . The patient refused the procedure.
>
> The patient's treating physicians at this point felt that she was mentally incompetent. Psychiatric and neurological consultations were requested to determine the possibility of dementia and/or mental incompetency. The psychiatric consultant felt that the patient was demented and not mentally competent to make decisions regarding her own care. This determination was based in large measure on the patient's steadfast "unreasonable" refusal to undergo surgery. The neurologist disagreed, finding no evidence of dementia. On questioning, the patient stated that she was refusing the hysterectomy because she *did not believe* she had cancer. "Anyone knows," she said, "that people with cancer are sick, feel bad and lose weight," while she felt quite well. The patient continued to hold this view despite the results of the biopsy and her physicians' persistent arguments to the contrary.

The physician seriously considered overriding the patient's refusal, because sound medical evidence demonstrated that she was unjustified in believing she

did not have cancer. As long as this patient continues to hold such a false belief and it is material to her decision, her refusal cannot correctly be said to be an informed refusal. Several complexities inherent in achieving effective communication are illustrated by this case: The patient was a poor white woman from Appalachia with a third-grade education. The fact that her treating physician was black turned out to be the major reason for her false belief that she had no cancer. She would not believe what a black physician told her. However, intense discussions with a white physician and with her daughter eventually resulted in a change in her belief and a consent to a successful hysterectomy.

The Problem of Waivers

A further problem about understanding is presented by waivers of informed consent. In the exercise of a waiver, a patient voluntarily relinquishes the right to an informed consent and relieves the physician from the obligation to obtain informed consent. The patient delegates decisionmaking authority to the physician or asks not to be informed. In effect, the patient makes a decision not to make an informed decision.

Some courts have held that "a medical doctor need not make disclosures of risks when the patient requests that he not be so informed,"[79] and some prominent writers in biomedical ethics hold that "rights are always waivable,"[80] including the right to an informed consent. Various studies indicate that perhaps sixty percent of patients want to know virtually nothing about certain procedures or the risks of those procedures, that a high percentage would consent without knowledge of risk, and that only a small percentage use the information provided in reaching their decisions.[81] Some physicians claim that more uninformed patients defer to physicians' recommendations than seek pertinent information, although one study also indicates that physicians tend to underestimate patient preferences for information.[82]

There are two primary ways to manage the problem of waivers. The contemplated medical procedure might be withheld until sufficient understanding is present, notwithstanding the patient's autonomously expressed desire not to be informed. Persons would then not be coerced or manipulated into receiving undesired information. Alternatively, if a patient or subject adequately understands his or her situation and then waives the right to relevant information about proposed medical interventions, the professional might proceed without insisting on any understanding beyond the person's understanding that he or she is waiving a right. In this second approach, the waiver constitutes a valid consent to therapy or research, even if it is not an informed consent.

It is usually appropriate to recognize waivers of rights because we have discretion over rights, and because waivers do seem justified in many contexts of consent. For example, if a committed Jehovah's Witness were to inform a doctor that he wished to have everything possible done for him, but did not want to know if transfusions or similar procedures would be employed, it is difficult to construct a moral argument (although legal reasons might exist) to support the conclusion that he must give a specific informed consent to the transfusions. Nevertheless, a general practice of allowing waivers is dangerous. Many patients have an inordinate trust in physicians, and the general acceptance of waivers of consent in research and therapeutic settings could make patients more vulnerable to those who would abbreviate or omit consent procedures for convenience, already a serious problem in health care. Accordingly, the danger of abuse of the waiver in busy medical settings, together with problems of how to determine the conditions under which a patient can make a voluntary and informed decision to waive the right to relevant information, demand caution in implementing waiver policies.

No general solution to these problems about waivers is likely to emerge. Each case or situation of waiver needs to be considered separately. There may, however, be appropriate procedural responses. For example, rules could be developed that disallow waivers except when they have been approved by deliberative bodies, such as institutional review committees and hospital ethics committees. If a committee determined that the person's interest was best protected in a particular case by recognizing a proposed waiver, it could be sustained. This procedural solution is not simply an evasion of the problem. It would be easy to violate autonomy and to fail to discharge our responsibilities by inflexible rules that either permit or prohibit waivers in institutional settings. Close monitoring by review could provide the necessary level of protection for patients, as well as a flexible arrangement for deliberation and decision.

Voluntariness

Autonomous persons typically consider the freedom to act as no less important than adequate understanding. Under the category *voluntariness,* we will concentrate on a person's independence from others' manipulative and coercive influences. As the law has long recognized, a consent or refusal coerced by threats or manipulated by misrepresentation is invalid.

Our use of the term *voluntariness* is intentionally narrow to distinguish it from broader uses that make it synonymous with autonomy. Some have analyzed voluntariness in terms of the presence of adequate knowledge, the absence of psychological compulsion, and the absence of external constraints.[83] If we adopted this broad meaning, the voluntariness condition would be the necessary and sufficient condition of autonomous action. However, we hold only that a person acts voluntarily to the degree he or she wills the action without being under the control of another influence. We consider here only control by other individuals. However, voluntariness can also be diminished or

voided by conditions such as debilitating disease, psychiatric disorders, and drug addiction.

Control over another person is necessarily an influence, but not all influences are controlling. If a physician orders a reluctant patient to undergo cardiac catheterization and coerces the patient into compliance through a threat of abandonment, then the patient is influenced by the physician's control. If, by contrast, a physician persuades the patient to undergo the procedure when the patient is at first reluctant to do so, then the patient is influenced by, but not controlled by the physician's actions. Many influences are resistible, and some are welcomed rather than resisted. The broad category of influence includes acts of love, threats, education, lies, manipulative suggestions, and emotional appeals, all of which can vary dramatically in their impact on persons.

Forms of Influence

Three primary categories of influence are present in our analysis: coercion, persuasion, and manipulation. Coercion, as we define it, occurs if and only if one person intentionally uses a credible and severe threat of harm or force to control another.[84] The threat of force or punishment used by some police, courts, and hospitals in acts of involuntary commitment for psychiatric treatment is a typical form of coercion. Society's use of compulsory vaccination laws is another. For a threat to be credible, both parties must believe that the person making the threat can effect it, or the one making the threat must successfully deceive the person threatened into so believing. A physician in a prison who tells an inmate he *must* submit to sedation will need an accompanying prison guard for the threat to be credible; only then will coercion occur.

Some threats will coerce virtually all persons (for example, a credible threat to kill the coerced person), whereas others will coerce only a few persons (for example, a threat presented by an employee to an employer of quitting a job unless a raise is offered). Whether coercion occurs depends on the subjective responses of the intended target of the coercion. However, a subjective response in which persons comply because they *feel* threatened does not qualify as coercion, because coercion requires that a real, credible, and intended threat be brought on a person so that his or her self-directedness is displaced. Coercion, so understood, voids an act of autonomy (that is, coercion renders even intentional and well-informed behavior nonautonomous), and so should be placed at one end of a continuum of types of influence.

In *persuasion,* as we use the term, a person must be convinced to believe in something through the merit of reasons advanced by another person. Accordingly, we do not recognize as a form of persuasion what Paul Appelbaum and Loren Roth label "forceful persuasion," which involves persistent forcefulness and sometimes misleading language. They cite a case of an intern who did not

accept a patient's refusal of an X-ray. The intern insisted that he "absolutely must have the film and that he could not refuse it." The patient then reluctantly agreed.[85] In our usage, neither nonrational nor forceful "persuasion" qualifies as a form of persuasion, because both are forms of manipulation.

The word *manipulation* is a generic label for several forms of influence that are neither persuasive nor coercive. The essence of manipulation is swaying people to do what the manipulator wants by means other than coercion or persuasion. For purposes of decisionmaking in health care, the key form of manipulation is informational manipulation, a deliberate act of managing information that nonpersuasively alters the person's understanding of a situation and thereby motivates him or her to do what the agent of influence intends. Many forms of informational manipulation are incompatible with autonomous decisionmaking. For example, deception that involves such strategies as lying, withholding information, and misleading exaggeration to cause persons to believe what is false are all inconsistent with autonomous choice. In Milgram's obedience experiments, he deceived subjects about virtually every aspect of the research to which they consented. By our criteria, these consents were manipulated and, therefore, did not qualify as autonomous choices.

Several problems encountered previously in discussing understanding reappear as issues of informational manipulation. An underdiscussed problem in health care is the amount of routine care and testing ordered in health care facilities without an explanation to the patient, thereby denying the patient a choice among alternatives (if any exist) and a right to refuse. Far more often mentioned is clinicians' uses of the therapeutic privilege to withhold information to manipulate patients into consenting to a medically desirable procedure.[86] The manner in which information is presented by tone of voice, by forceful gesture, and by framing information positively ("we succeed most of the time with this therapy") rather than negatively ("we fail with this therapy in thirty-five percent of the cases") can easily manipulate a patient's perception and response, and thereby affect understanding. The major concern about manipulation in medicine has been expressed by Eliot Freidson:

> It is my impression that clients are more often bullied than informed into consent, their resistance weakened in part by their desire for the general service if not the specific procedure, in part by the oppressive setting they find themselves in, and in part by the calculated intimidation, restriction of information, and covert threats of rejection by the professional staff itself.[87]

Nevertheless, one can easily inflate the threat of control by manipulation beyond its significance in health care. We typically make decisions in a context of competing influences, such as personal desires, familial constraints, legal obligations, and institutional pressures. Although significant, these influences need not be controlling to a substantial degree. From the perspective of deci-

sionmaking by patients and subjects, we need only establish general criteria for the point at which autonomous choice is imperiled, while recognizing that no sharp boundary can be drawn in many cases between controlling and noncontrolling influences. Also, in each case the powers of an individual patient or subject must be assessed, and the health professional will need to consider the particular patient's subjective resistibility to influence, not the (so-called objective) reasonable person's ability to resist.

The Obligation to Abstain from Controlling Influence

Thus far we have attempted primarily to distinguish influences that are compatible with substantial autonomy from influences that are not compatible. Now we can examine the justifiability of exerting these forms of influence.

Many influences are welcomed by patients and subjects, and even unwelcome influences are sometimes compatible with autonomous decisionmaking. In some cases professionals are morally blameworthy if they do *not* attempt to persuade resistant patients to pursue treatments that are medically essential, and such persuasion need not violate respect for autonomy. Reasoned argument in defense of an option is a form of providing information and is often vital to ensuring understanding. It is never an unjustified form of influence, although in some cases it can be unduly intrusive and therefore unjustifiable.

We are assuming that influence by appeal to reason—persuasion—is in theory and practice distinguishable from influence by appeal to emotion. As applied to health care professionals, the problem is to distinguish emotional responses from cognitive responses and to determine which are likely to be evoked. The goal is to avoid overwhelming the person with frightening information, particularly if the person is in a psychologically vulnerable or compromised state. Disclosures or approaches that might rationally persuade one patient might overwhelm another patient whose fear or panic would short-circuit reason.

Coercion and controlling manipulation are occasionally justified, although these occasions are infrequent in medicine (by contrast to police work, where such techniques are more common and also more commonly justified). If a physician responsible for a disruptive and childishly noncompliant patient threatens to discontinue treatment unless the patient alters certain behaviors, the physician's mandate may be justified even though coercive. The most difficult problems about manipulation concern not punishment and threat, which are almost always unjustified in health care and research, but rather the effect of rewards, offers, and encouragement. An egregious example of an unjustified offer occurred during the aforementioned Tuskegee Syphilis experiments. Various offers were used to stimulate and sustain the interest of subjects in continued participation. Subjects were offered free burial assistance and insurance, free transportation to and from the examinations, and a free stop in town on the return trip. They were rewarded with free medicines and free hot meals on the days of the examination. The socioeconomic deprivation of these subjects made them vulnerable to these overt and unjustifiable forms of manipulation.[88] Problematic techniques in clinical practice are usually far more subtle and difficult to locate and analyze, but they can have the identical effect.

When an offer is made in a setting in which it is abnormally attractive—for example, an offer of large sums of money or freedom for destitute prisoners—it may be manipulative, but it is never coercive. To maintain that irresistibly attractive offers such as free medical care or freedom from involuntary commitment coerce patients or subjects deeply distorts the concept of coercion (unless the "offer" is in truth a disguised threat), because then anyone who intentionally and successfully influenced another by presenting an offer so attractive that the person was unable to resist it—for example, a large salary at a wonderful job—would have coerced the person. An offer of something irresistible is not coercive, although it does under some conditions manipulatively take advantage of a person's vulnerabilities.

The conditions under which an influence is controlling and morally unjustified may be clear in theory, but they are often unclear in concrete situations, and many borderline cases remain. Some difficult cases in health care consist of manipulation-like situations in which patients or subjects are in desperate need. To say that a person desperately needs something, such as a medication or a source of income, means that without it a high probability exists that the person (or some loved one) will be seriously harmed. Attractive offers such as free medication or extra money can leave persons without any meaningful choice besides accepting the offer. Such a person is constrained in a desperate situation, but not controlled by another person's intentional manipulation.

Some seem to believe that an offer of this magnitude to a person in desperate need is inherently exploitative. In some circumstances the offer is likely to appear to the beneficiary as a threat—for example, if an experimental therapy is the sole therapy and can be obtained only if a person becomes a research subject. However, such an offer is sometimes perceived differently. In 1722, Newgate Prison officials offered several inmates their freedom, as an alternative to hanging, if they volunteered to be subjects in an experiment on smallpox inoculation.[89] It might at first seem that they were coerced, because the offer appears to be a disguised threat on the order of "We will hang you unless you become an experimental subject." More plausibly, though, this manipulation-like circumstance involves a welcome offer made to persons in desperate need, who without the offer would be hanged anyway. The prisoners certainly considered the offer to be fortuitous, as these condemned men all survived and were released.

In contrast, influences that ordinarily are resistible can become controlling

for abnormally weak, dependent, and surrender-prone patients, and compliance may be induced in these patients by contributing to or playing on their desperation, anxiety, boredom, or other emotions. The hope of more attention and better care can be a significant factor for a bedridden person. What a health professional intends as an attempt at rational persuasion can irrationally influence the patient by attacking his or her vulnerabilities. We are not implying that health professionals do routinely manipulate or exploit patient vulnerabilities, but only that many patients are susceptible to this kind of influence and need protection against it.[90]

It is especially vital to ensure that conditions permitting resistance to control are preserved in total institutions, whose populations are admitted involuntarily. The threat of exploitation is substantial in these institutions, yet neither coercive institutionalization nor coercive institutions entail that each decision made by a person in the institution is coerced. There is no reason why prisoners, for example, cannot validly consent to some research if coercive tactics are not involved in enlisting them as subjects and if there are no manipulative offers, such as unduly large payments for excessive risk taking.[91]

These problems are often more subtle and difficult in institutions to which persons are admitted voluntarily but in which rules, policies, and practices can work nonetheless to compromise autonomous choice. Perhaps nowhere is this compromise more evident than in long-term care. For example, the elderly in nursing homes frequently experience a constriction of their choices, particularly in routine or everyday matters. Many people in nursing homes have already suffered a decline in their ability to carry out personal choices because of physical impairments. This decline in *executional* autonomy need not be accompanied by a decline in *decisional* autonomy, and yet their autonomous choices and decisions are often neglected or overridden by the nursing home.[92]

These everyday matters often range over food (when, what kind, how prepared, and how much), roommates (who selects them and how to resolve conflicts), possessions (which to keep and how to protect them), exercise (when, what kind, and with what supervision), sleep (when and how much), clothes (what to wear and when to wash), as well as baths, medications, and restraints. The liberty of competent residents to live their lives in accord with their preferences and life plans must often be balanced against protecting their health, protecting the interests of others, promoting safety and efficiency in the facility, and allocating limited financial and other resources. Although respect for autonomy suggests individualized care in the nursing home setting, such care can rarely be individualized in the ways we expect outside such institutions.

Consider the following example from a facility named Mansion Manor. Mrs. Hollinger, who is seventy-six years old, has encountered difficulty with the nurses's aides about the facility's requirement that residents rise in time for breakfast, which is served at 7:30. She has never liked breakfast and she moves slowly after a second stroke. She finds the effort to make it to breakfast almost intolerable. She has been late each morning for two weeks, and the aides say that her late arrival disrupts the feeding of other residents and that she fails to finish her breakfast when she is late because she is a slow eater. The floor nurse warns her, in a manner she finds threatening, that if her tardiness for breakfast continues, she will be put with those who cannot feed themselves in a separate dining room. Her worries about being late for breakfast have begun to cause her trouble in sleeping and have increased her tiredness. The staff has had several earlier battles with Mrs. Hollinger, particularly ones centered on her charges that the staff "poked around" in her room and removed some of her belongings that were deemed to be unsafe to others who might wander into her room. The staff's animosity makes it difficult to determine whether Mrs. Hollinger's late arrival at breakfast in fact infringes on the rights of others and disrupts institutional order.

The staff's efforts at persuasion are justifiable, but their coercive threat to put her in the separate dining room is not justified unless her actions genuinely pose problems for others or for the institution. The aides could respond that they are not overriding Mrs. Hollinger's autonomy, only respecting it, because she accepted the rules and regulations that restrict liberty when she voluntarily entered the nursing home. Thus, the argument might go, she has an obligation of compliance, not only because of the need for institutional order but also because of her consent. However, before this claim can be sustained, we would need to know exactly what Mrs. Hollinger (and her son) were told at the outset about the rules and regulations. They might well have grounds to complain about initial disclosures or about the nursing home's narrowly legalistic interpretations of its rules and of governmental regulations.

The director of nursing explained that federal regulations require that the first meal of the day occur no more than 14 hours after a substantial evening meal the previous day, which for some residents is 4:30 p.m. This regulation is indispensable to protect nursing home residents from exploitation and harm. However, it could also be construed as establishing an option right, rather than a mandatory right, for residents. An option right can be waived, whereas a mandatory right such as the right to education cannot.[93] If an option right exists, autonomous residents would have the option to accept or refuse the meals. A strict interpretation of a mandatory right would require force-feeding of resistant autonomous residents, and the implausibility of such an interpretation provides a reason to suspect that the institution is displaying bad faith in its conflict with Mrs. Hollinger. Even if the principle of respect for autonomy can be justifiably overridden to protect others or to establish (legitimate) institutional order, the institution must choose the least restrictive alternative.[94]

Some contend that respect for autonomy, taken literally, demands too much of nursing homes and other long-term care facilities. They propose that this

individualistic principle be replaced by a communitarian perspective in which informed consent is superseded by "negotiated consent" and individual rights are incorporated into a larger vision of community, with an emphasis on mutual responsibilities.[95] Although this alternative has attractive features, it is too obscure and risky without explicit protections against violations of autonomy. The presumption in favor of voluntariness and rights of autonomous choice should never be renounced. However, in some cases other persons—surrogate decisionmakers—should be granted some measure of decisional authority for residents, for reasons now to be discussed.

A Framework of Standards for Surrogate Decisionmaking

Surrogate decisionmakers reach decisions for doubtfully autonomous or nonautonomous patients. If a patient is not competent to choose or to refuse treatment, a hospital, a physician, or a family member may justifiably be placed in a decisionmaking role or go before a court or other authority to seek resolution of the issues before a decision is implemented. Courts and legislatures have been actively involved in this area since the Quinlan decision in 1976, and significant advances have been made in both law and ethics. However, much remains undecided, particularly with regard to patients who are incompetent and debilitated, yet conscious. Many judgments about terminating or continuing treatment are made daily for patients in this condition—for example, those suffering from stroke, Alzheimer's disease, Parkinson's disease, chronic depression affecting cognitive function, senility, and psychosis.

Celebrated legal cases have centered on formerly autonomous patients, including Karen Ann Quinlan, Earle Spring, Brother Fox, Claire Conroy, Paul Brophy, and Nancy Cruzan, as well as never-competent patients such as Joseph Saikewicz and John Storar. In such cases, courts have split over the use of two surrogate decisionmaking standards: best interests and substituted judgment. Neither has an entirely clear basis in autonomy, but appeals to autonomy are often used by defenders of these standards. Currently a received opinion operates in many courts about how treatment decisions should be reached for both formerly competent and never-competent patients. In this account, all patients have a right to decide, and their autonomous choices must be consulted whenever possible as the basis of any decision; an incompetent person is still a person with a right to choose. We will resist this framework, substituting a different account of decisionmaking standards and of the order of priority among them.

We will consider three general standards that surrogate decisionmakers might use: *substituted judgment*, which is often presented as an autonomy-based standard, *pure autonomy*, and the *patient's best interests*. Our objective is to structure and integrate this framework of standards for surrogate decisionmaking.

Although we assess these standards for law and policy, our underlying argument is independent of law and policy. The argument is a moral one that extends our earlier discussions of the value of protecting autonomy. Only in Chapter 4 will we consider *who* should be the surrogate decisionmaker.

The Substituted Judgment Standard

The standard of substituted judgment appears initially to be autonomy-based, and several influential judicial opinions have so viewed it. However, it is, at best, a weak autonomy standard. Substituted judgment begins with the premise that decisions about treatment properly belong to the incompetent or nonautonomous patient by virtue of rights of autonomy and privacy. The patient has the right to decide but is incompetent to exercise it. It would be unfair to deprive an incompetent patient of decisionmaking rights merely because he or she is no longer (or has never been) autonomous. Nonetheless, another decisionmaker should be substituted if the patient is currently unable to make autonomous decisions.

This standard requires the surrogate decisionmaker to "don the mental mantle of the incompetent," as the *Saikewicz* court put it—that is, to make the decision the incompetent would have made if competent. In *Saikewicz*, the court had to consider evidence that most reasonable persons with Joseph Saikewicz's illness choose treatment, but the court invoked the standard of substituted judgment to decide that Saikewicz, a never-competent patient, would not have chosen treatment had he been competent. The court defined its task as determining "how the right of an incompetent person to decline treatment might best be exercised so as to give the fullest possible expression to the character and circumstances of that individual." Asserting that what the majority of reasonable people would choose could differ from what a particular incompetent person would choose, the court proposed the following standard:

> [T]he decision in many cases such as this should be that which would be made by the incompetent person, if that person were competent, but taking into account the present and future incompetency of the individual as one of the factors which would necessarily enter into the decision-making process of the competent person.[96]

Both the *Quinlan* and the *Saikewicz* courts used the substituted judgment standard, first attempting to determine the individual's subjective wants and needs and then attempting to decide how to proceed in light of the individual's value system. However, these two cases involve different interpretations and uses of the substituted judgment standard. In *Quinlan*, the court sought to protect an autonomy right for a person who could not assert the right because of her permanent vegetative state. The court authorized the patient's father to infer her wants and needs from her life as a competent person, despite the fact that

his judgments involved several assumptions. The court did not frame the issue either in terms of parental rights over children or in terms of whether the father could decide in her best interests. Rather, the court attempted to protect her autonomy and privacy rights by asking the father to determine what she would have chosen if she had been able to choose. The court found that "Karen's right of privacy may be asserted on her behalf by her guardian whether she would exercise her right" to terminate care, "even if it meant the prospect of natural death."[97] This decision illustrates the scope courts often give to rights of privacy and autonomy.

In *Saikewicz*, the lack of evidence about the incompetent patient's likely choice forced the court to look to what is known about other people in his situation to help determine what a reasonable person in his circumstances, with his needs and desires (insofar as they are ascertainable), would decide. As understood in *Saikewicz*, and to some extent in *Quinlan*, the premise of the substituted judgment standard has a fictional component. An incompetent person cannot literally be said to have the right to make medical decisions if the right can only be exercised by competent persons. This fictional quality makes substituted judgment controversial. John Robertson has argued that it is desirable to treat incompetent persons as autonomous, despite the apparent absurdity of treating them in a way that diverges from their situation: "Eliminating this divergence would mean that we treat the incompetent in all respects as a non-thinking, nonchoosing, irrational being—in short, as a nonperson."[98]

Despite its established status, the standard of substituted judgment should be used for once-competent patients only if reason exists to believe that a decision can be made as the patient would have made it. In such cases, the surrogate's acquaintance with the patient should be sufficiently deep and relevant that a judgment will reflect the patient's goals and views. If the surrogate can reliably answer the question, "What would the patient want in this circumstance?," then substituted judgment is an appropriate standard. But if the surrogate can only answer the question, "What do you want for the patient?" then this standard is inappropriate, because all connection to the patient's former autonomy has vanished.

Similarly, the standard of substituted judgment should be rejected for never-competent patients, because their autonomy is not involved. No basis exists for a judgment of autonomous choice if a person has never had autonomy. The never-competent patient therefore should be considered relevantly different from those who can now make or who previously have made autonomous choices. Exponents of substituted judgment have failed to establish the relevance of the characteristic of autonomy for never-autonomous patients, leading one court to hold that trying to determine what a never-competent patient would have decided if competent is like asking, "If it snowed all summer, would it then be winter?"[99]

There are also problems with substituted judgment when used for incompetent but conscious patients such as Earle N. Spring, a senile man whose family and physicians considered his continuation on kidney dialysis to be of doubtful value (see Case 5 in the appendix). His wife and son petitioned a Massachusetts court for authorization to terminate dialysis. If a surrogate views such a patient in terms of what the patient might wish if competent, rather than in terms of his medical need, a danger arises of overlooking the person's stake in continued existence. Many debilitated incompetent patients, like Earle Spring, have their care terminated on the basis of highly tenuous judgments by financially strained relatives about what the person would have wanted if he or she could speak. In general, little is known at present about how accurately surrogate decisionmakers reflect the preferences of patients.[100] In the instance of *Spring*, the Massachusetts Supreme Court (unlike the original probate court) accepted the family's argument that Earle Spring had been an active, vigorous outdoorsman who hated confinement and killed suffering animals when he found them in the woods. Almost all senile individuals suffer a loss of activity, but this has never been considered a good or sufficient reason for terminating their lives.

Preservation of privacy and dignity are often given as reasons in such cases, even when for months little privacy or dignity has been available for such patients. A best interests test would lead to closer scrutiny of questions about the patient's welfare than does substituted judgment. This is a matter of practical importance, because many residents of nursing homes and state facilities routinely have such judgments made about their lives—especially regarding continuation of respirators, antibiotics, nutrition and hydration, and the like. We do not suggest that these judgments are poorly made, but we do hold that substituted judgment is a poor basis on which to make them. (There are parallel dangers with authorizing parental decisions for healthy minors—for example, when a parent makes a "substituted judgment" about whether a healthy minor child would want to give a kidney to a sibling.)

The rule of substituted judgment, then, helps us understand what should be done for once-competent patients whose relevant prior preferences can be discerned; but, so interpreted, it collapses into a pure autonomy standard that respects previous autonomous choices. We conclude that we should abandon substituted judgment insofar as possible in law and in ethics and substitute a pure autonomy standard in contexts in which explicit prior autonomous judgments are identifiable.

The Pure Autonomy Standard

The second standard, then, eliminates the ghost autonomy found in substituted judgment. It applies exclusively to formerly autonomous patients who expressed a relevant autonomous decision or preference. This standard makes

more specific the general commitments of the principle of respect for autonomy. One must respect past, self-regarding, autonomous decisions reached by now-incompetent but previously competent persons. Whether or not there exists a formal advance directive, prior autonomous judgments should be accepted. (It is assumed that such judgments are known, not merely conjectured, and are directly relevant to a contemplated action.) We find an instructive approach to such formerly autonomous patients in the Claire Conroy case, in which the New Jersey Supreme Court grappled with several standards of surrogate decisionmaking.[101]

Claire Conroy, an eighty-three-year-old nursing home resident, suffered from irreversible physical and mental impairments, including organic brain syndrome, arteriosclerotic heart disease, hypertension, diabetes, necrotic ulcers on her left foot, and a gangrenous left leg. She was awake enough to track persons with her eyes, but was severely demented, lay in a fetal position, and was unable to speak. She had no discernible cognitive or volitional functioning. She could not swallow enough food and water to sustain herself, and she received nutrition and hydration through a nasogastric tube. She could move a little, but could not control her excretory functions. Certain stimuli resulted in an occasional response. For example, she would sometimes smile when her hair was combed or when she received a comforting rub, and she would occasionally moan when moved or fed or when her bandages were changed.

Claire Conroy's nephew (Thomas Whittemore) was her guardian and only surviving blood relative. He sought court permission to remove his aunt's nasogastric tube, which would result in her dehydration and death in about a week. His petition was opposed by her physician, who viewed such an action as a violation of medical ethics. The trial court decided to permit removal of the feeding tube, although her dying might be painful, on grounds that her life had become permanently burdensome. The court ordered removal, but a court-appointed guardian ad litem appealed, and the order was stayed pending appeal. Conroy died during the appellate process, but two courts nevertheless issued opinions.

A first appellate court reversed the trial court's judgment, on grounds that removal of the feeding tube would cause her death and thus would constitute an active and impermissible killing from dehydration and starvation. On further appeal, the New Jersey Supreme Court held that any medical treatment, including artificial nutrition and hydration, may be withheld or withdrawn from an incompetent patient under some circumstances. The court invoked the incompetent patient's autonomy right to accept or refuse medical treatment, even though the right must be exercised by another decisionmaker. This court's language appears at first to be a mainline instance of the substituted judgment standard. It asserts that "the goal of decision-making . . . should be to determine and effectuate . . . the decision that the patient would have made if competent" and that "the right of an adult who, like Claire Conroy, was once competent, to determine the course of her medical treatment remains intact even when she is no longer able to assert that right or to appreciate its effectuation."

However, this court also makes direct appeals both to a pure autonomy standard and to a best interests standard. The court holds that life-sustaining treatment is legitimately withheld or withdrawn from an incompetent patient when it is clear from a "subjective test"—a demonstrable basis in former autonomous choices—that this particular patient, when autonomous, would have refused under the circumstances. If this subjective or autonomy test is not met, a best interests test then must be satisfied. The court here recognizes the patient's right of informed refusal as correlative to the right of informed consent.[102] The court reasons that the subjective or autonomy-based standard is in principle fulfilled by a written document (such as a living will); an oral directive to family member, friend, or health care provider; a durable power of attorney; the patient's convictions about medical treatment administered to others; religious beliefs and tenets; or the "patient's consistent pattern of conduct with respect to prior decisions about his own medical care."

The court indicates that it had erred a decade earlier in *Quinlan* when it disregarded the evidence of "statements that Ms. Quinlan made to friends about the artificial prolongation of the lives of others who were terminally ill." Such evidence is "certainly relevant." But, the court notes, evidence has different degrees of probative value, "depending on the remoteness, consistency, and thoughtfulness of the prior statements or actions and the maturity of the person at the time of the statements or acts." For example, advance directives typically express a person's general standards, designate a decisionmaker, or combine the two. A living will that specifies personal standards for decisionmaking—for example, "I don't want to be kept alive by a respirator if I am permanently comatose"—can have high probative value, though it requires interpretation in particular situations. The court reasons that recognizing such advance directives respects incompetent patients' (previous) autonomy.

Ideally, a surrogate conveys rather than substitutes another's autonomous judgment. In many cases, however, questions exist about the reliability of evidence used to determine the patient's earlier preferences, such as whether he or she was sufficiently competent and clearly expressed relevant preferences. *Conroy* rightly notes that "in the absence of adequate proof of the patient's wishes, it is naive to pretend that the right to self-determination serves as the basis for substituted decision-making." (In the later, landmark *Cruzan* decision, the U.S. Supreme Court held that a state can legitimately require clear and convincing evidence of a patient's prior express wishes about forgoing or withdrawing life-sustaining treatment, rather than accepting a proxy's interpretation of those wishes.) Nevertheless, the *Conroy* court holds that an absence

of adequate proof does not entail that life-sustaining treatment must be continued: "Life-sustaining treatment may also be withheld or withdrawn from a patient in Claire Conroy's situation if either of two types of 'best interest' test—a *limited-objective* or a *pure-objective* test—is satisfied."

In its reliance on "best interests," the court does not substantially deviate from considerations of autonomy in its "limited-objective test." Its test requires "some trustworthy evidence that the patient would have refused the treatment" along with the decisionmaker's conviction that "the burdens of the patient's continued life with the treatment outweigh the benefits of that life for him." If these conditions are satisfied, the treatment is deemed merely to prolong suffering. Any evidence mentioned under the subjective (pure autonomy) test could be sufficient in reaching judgments about the relevant burdens, although it might be "too vague, casual, or remote to constitute the clear proof of the patient's subjective intent that is necessary to satisfy the subjective test."

Even if no evidence exists about the patient's previous wishes, life-sustaining treatment is justifiably withheld or withdrawn if decisionmakers satisfy the pure-objective test, which is strictly a test of best interests: "The net burdens of the patient's life with the treatment should clearly and markedly outweigh the benefits that the patient derives from life" and "the recurring, unavoidable and severe pain of the patient's life with the treatment would be such that the effect of administering life-sustaining treatment would be inhumane." Although the New Jersey court limited its holding to previously competent patients in Claire Conroy's situation, its arguments about best interests extend to other classes of incompetent patients.[103] Paradoxically, the court determined that Claire Conroy herself did not satisfy any of the court's standards for withdrawing life-sustaining treatment, and, therefore, had she lived, the court would not have authorized removal of her feeding tube.

Although we also commend a pure autonomy standard, as appropriate, there are additional problems in *Conroy* and similar legal decisions regarding satisfactory evidence for acting under this standard. In the absence of explicit instructions, a surrogate decisionmaker might, for example, selectively choose from the patient's life history those values that accord with the proxy's own values, and then use only those selected values in reaching decisions. The proxy's findings might also be based on values of the patient that are only distantly relevant to the immediate decision—such as the patient's expressed dislike of hospitals. It is reasonable to ask what a decisionmaker can legitimately infer from Claire Conroy's prior conduct, especially her fear and avoidance of doctors and her earlier refusal to consent to amputation of a gangrenous leg.

A troublesome problem is that surrogates often assume an explicitness in a patient's directive about the future that does not, with sufficient directness, apply to the decision at hand. In *Evans v. Bellevue Hospital*, a formerly competent patient had executed a durable power of attorney authorizing another to make medical decisions in a circumstance of incompetence and had executed a second document refusing life-sustaining treatments if he suffered from "illness, disease or injury or experienced extreme mental deterioration, such that there is no reasonable expectation of recovering or regaining a meaningful quality of life." When the patient became incompetent and suffered from brain lesions due to toxoplasmosis, a form of infection, the designated surrogate refused treatment, allegedly following the executed document's declaration. Both physicians and a court rightly refused to recognize the proxy's decision, because the document did not clearly pertain to this condition, which was in principle treatable and had a chance of restoring the patient's capacity to communicate.[104] Such imprecise statements provide too little guidance and are sometimes dangerous. Often these cases need to be handled under the best interests standard rather than an autonomy standard, even when legally valid documents have been executed with the intent of exercising autonomous control.

There is also a procedural problem of ensuring that surrogates respect a patient's prior autonomous judgments or otherwise act responsibly as surrogates. It has become increasingly difficult to find suitable persons willing to assume the burdensome job of guardianship for institutionalized mentally disabled persons, and families sometimes make decisions that conflict with the apparent wishes of a now incompetent person. One study focused on decisions by surrogates (largely sons and daughters) for 168 elderly patients in nursing homes about whether to permit their participation as research subjects in a minimal risk study of morbidity associated with long-term urinary catheters. These surrogates tended to believe that research should not be conducted in nursing homes, that they themselves would not consent to participate, that the research would disturb the patient, and that the patient, if competent, would not consent. Nonetheless, fifty-four percent consented to the patient's participation in the study, and thirty-one percent of the surrogates who thought the patients would not consent if competent still consented for the patient. Because this discrepancy emerged only through interviews after completion of the project, the researchers did not confront the ethical dilemma of what to do when surrogates act against what they believe to be the patient's wishes.[105] The study's authors suggest that consent auditors are sometimes needed to ensure better decisions when the surrogate appears to act against the patient's preferences.

Another procedural problem has emerged from a recent study of gender bias in appellate judicial opinions regarding the termination of life-sustaining treatment for newly incompetent patients. The investigators concluded that "Judicial reasoning about profoundly ill, incompetent men accepts evidence of mens' treatment preferences to define the standing of personal autonomy in decisions about life-sustaining treatment. Judicial reasoning about women defines the role

of caregivers in making treatment decisions after either rejecting or failing to consider evidence of women's preferences with regard to life-sustaining treatment."[106] For newly incompetent patients without a written advance directive, the appellate courts tended to adopt the male patient's preferences from reports of family and friends, whereas they rarely took this approach for newly incompetent female patients. The following tendencies were discovered in this study: A man's prior opinions are typically viewed as rational, whereas a woman's earlier comments are often viewed as unreflective, emotional, or immature. A woman's viewpoint, as reflected in prior statements, is sometimes neglected altogether. Statements about women's views and values are also subjected to a higher standard of clear and convincing evidence. Finally, the court opinions tend to depict men as subject to medical assault and women as vulnerable to medical neglect. Both men and women are placed at risk by such gender bias. Men are vulnerable to quick decisions about termination of life-sustaining treatment, whereas women are at risk of not having their autonomous decisions taken seriously.

In summary, we have argued that previously competent patients who autonomously expressed their preference in the form of an advance directive should be treated under the pure autonomy standard, and we have suggested an economy of standards. It is presently popular in biomedical ethics to hold that an ordered set of standards for surrogate decisionmaking runs from (1) autonomously executed advance directives to (2) substituted judgment to (3) best interests, with (1) having priority over (2) and (1) and (2) having priority over (3) in a circumstance of conflict. We have collapsed (1) and (2) as essentially identical. Their defense and only basis is in the principle of respect for autonomy, which applies if and only if a relevant autonomous judgment exists that constitutes an authorization. Where the previously competent person left no reliable traces of his or her wishes, surrogate decisionmakers should adhere only to (3). This conclusion takes us to an examination of the best interests standard.

The Best Interests Standard

Under the best interests standard a surrogate decisionmaker must determine the highest benefit among the available options, assigning different weights to interests the patient has in each option and discounting or subtracting inherent risks or costs. The term *best* is used because the obligation is to maximize benefit through a comparative assessment that locates the highest net benefit. The best interests standard protects another's well-being by assessing risks and benefits of various treatments and alternatives to treatment, by considering pain and suffering, and by evaluating restoration or loss of functioning. It is therefore inescapably a quality-of-life criterion.

Although a best interests judgment must evaluate risks and benefits for the person involved, such a judgment should not rest solely on known subjective preferences or other forms of personal value. It appeals indirectly to these autonomy considerations insofar as they provide a basis for understanding welfare and interpreting interests. Autonomous preferences should be considered under the best interests standard only as far as they affect interpretations of quality of life, direct benefit, and the like. They should be known preferences, not inferences about what the patient would think based on his or her broad expressed values.

The best interests standard has been widely used both in and beyond health care settings. Long before autonomy and privacy were pervasively applied through law to incompetents and minors, the responsibility of parents toward their children was legally defined as the responsibility to act in the best interests of those children. It was assumed in law that parents generally do act in their children's best interests and that the state should not interfere except in extreme circumstances in which the state and the parents disagree about some decision with potentially serious consequences for the child—for example, when Jehovah's Witness parents refuse lifesaving blood transfusions for their minor children. If a court rather than the family decides, the court has already made a judgment about the unjustifiability of the family's proposed course of action (or the family's incompetence to decide).

We believe there are circumstances in which the best interests standard, so understood, can validly be invoked to override advance directives executed by autonomous patients who have become incompetent, refusals by minors, and refusals by mental patients. This overriding can occur, for example, in a case in which a person has designated another by a durable power of attorney to make medical decisions on his or her behalf. If the designated surrogate makes a decision that is clearly against the patient's best interests, the decision should be overridden unless there is a clearly worded, second document executed by the patient that specifically supports the surrogate's decision. Overriding the surrogate's decision in such a case does not violate respect for autonomy or constitute a paternalistic intervention.

For previously competent patients whose prior preferences cannot be reliably traced and for never-competent patients, it is appropriate to rely on a best interests standard as more suitable than the pure autonomy standard and substituted judgment standards. Although it has been argued that substituted judgment is a standard that seeks "to implement the patient's best interests as that patient would have defined them" and, therefore, that "the substituted judgment approach is merely one way in which the best interests standard is given content,"[107] this synthesis is too convenient and confusing. The best interests standard can in principle conflict with either an autonomy standard or a substituted judgment standard. As the authors of the above-mentioned study of surrogate

decisions for the elderly note, these standards do in fact conflict in many cases. It is best to keep these standards as conceptually and normatively distinct as possible.

Courts, health care institutions, and religious traditions have too long been eager to assert that they do not make quality of life judgments, but only reach decisions in view of what the patient would have chosen. The substituted judgment standard has been popular, because it enables a decisionmaker to disclaim quality-of-life considerations altogether, while claiming to look exclusively at the individual's preferences. Courts have viewed quality-of-life judgments as comparative ways of expressing a person's social worth, and they have understandably wanted to avoid comparative ranking of the worth of individual lives. However, "quality-of-life judgments" are not about the social worth of individuals, but about the value of the life for the person who must live it. The value of a life is primarily (although not exclusively) the value it has for that person. Best interests judgments are one way to focus attention on this point, rather than on the value the person's life has for other persons. Accepting a best interests standard, properly so called, is tantamount to acknowledging that we have to decide in marginal cases what a patient's welfare interests are at the moment, not what they would have chosen in some imaginary possible world.

Unfortunately, the best interests standard has sometimes been interpreted as highly malleable, permitting values that are irrelevant to the patient's benefits or burdens and incorporating intangible factors of questionable value to the incompetent person. For example, when parents have sought court permission for a kidney transplant from an incompetent minor child to a competent sibling, parental judgments about the "donor's" best interests have on occasion taken into account projected psychological trauma from the death of the sibling and the psychological benefits of the unselfish act of "donation."[108] While we would not exclude such considerations altogether, they should be greeted with skepticism and with additional procedural protections, such as committee review. Best interests judgments should concentrate on tangible factors, such as physical suffering and medical diagnosis, and should be extended into other domains only with hesitation and great caution.

Under this formulation, questions arise about whether the burdens considered under the best interests standard should be limited to physical pain and suffering, as judicial language often suggests. If pain and suffering were the only relevant burdens, it would be difficult to justify withholding or withdrawing life-sustaining treatment for a permanently comatose patient. However, this range of concerns about the best interests standard cannot be examined until Chapters 4 and 5, where we discuss benefits and harms more comprehensively.

Conclusion

The intimate connection between autonomy and decisionmaking in health care unifies this chapter's several sections. Although we have justified the obligation to solicit decisions from patients by the principle of respect for autonomy, we have acknowledged that the principle's precise demands remain unsettled and open to interpretation and specification. For example, notable issues about the principle's connection to rules of truthfulness, confidentiality, and privacy are discussed in Chapter 7. We have thus far argued only that making respect for autonomy a trump moral principle, rather than one moral principle in a system of principles, gives it an excessive value. The human moral community, indeed morality itself, is rooted no less deeply in the three clusters of principles discussed in subsequent chapters. In many clinical circumstances the weight of respect for autonomy is minimal, and the weight of nonmaleficence or beneficence is maximal. Similarly in public policy, the demands of justice can easily outweigh the demands of respect for autonomy.

Several conclusions in this chapter could be viewed as one-sided in their deference to autonomy, on grounds that their implications for professional conduct exceed current legal and regulatory requirements. For example, our proposed shift from a focus on disclosure to a focus on understanding and effective communication entails a different and more burdensome way of structuring the consent-solicitation process. Nevertheless, we have not argued that our proposals should be transformed directly into enforceable legal requirements or into regulatory rules or hospital policies. It is not always justifiable to provide the resources needed to create a context in which professionals conduct themselves in accordance with the full set of strategies suggested in this chapter. This problem invites judgments about justice and resource allocation (see Chapter 6) that compete with the obligation to obtain informed consent.

Notes

1. The core idea of autonomy has been helpfully treated by Isaiah Berlin, "Two Concepts of Liberty," in *Four Essays on Liberty* (Oxford: Oxford University Press, 1969), pp. 118–72; Joel Feinberg, *Harm to Self*, vol. III in *The Moral Limits of Criminal Law* (New York: Oxford University Press, 1986), ch. 18 and 19; and Thomas E. Hill, Jr., *Autonomy and Self-Respect* (Cambridge: Cambridge University Press, 1991), ch. 1–4.
2. See Gerald Dworkin, *The Theory and Practice of Autonomy* (New York: Cambridge University Press, 1988), ch. 1–4; Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5–20.
3. See Stanley Benn, "Freedom, Autonomy and the Concept of a Person," *Proceed-*

ings of the Aristotelian Society 76 (1976): 123–30. In his earlier and later views, Benn notes that one can and should be a proper object of respect without satisfying the "exacting requirements of the ideal of autonomy." See *A Theory of Freedom* (Cambridge: Cambridge University Press, 1988), pp. 3–6, 155f, 175–83. For an emphasis on autonomy through authorship of one's life, see Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), pp. 145–62, 368–429, esp. 154–56, 368–72.

4. Dworkin, *The Theory and Practice of Autonomy*, pp. 15–20.
5. For practical implications and empirical studies, see Priscilla Alderson, "Consent to Children's Surgery and Intensive Medical Treatment," *Journal of Law and Society* 17 (1990): 52–65; and Barbara Stanley et al., "The Functional Competency of Elderly at Risk," *The Gerontologist* 28, Suppl. (1988): 53–58.
6. See Robert Paul Wolff, *In Defense of Anarchism* (New York: Harper and Row, 1970), pp. 4–6, 13f, and Arthur Kuflik, "The Inalienability of Autonomy," *Philosophy and Public Affairs* 13 (1984): 271–98. See also Joseph Raz, "Authority and Justification," *Philosophy and Public Affairs* 14 (1985): 3–29; and Christopher McMahon, "Autonomy and Authority," *Philosophy and Public Affairs* 16 (1987): 303–28.
7. Susan Sherwin, *No Longer Patient: Feminist Ethics and Health Care* (Philadelphia: Temple University Press, 1992), p. 138. However, she does not reject the relevance of moral considerations of autonomy. For a view that "feminists have reason to regard institutions and practices that undermine autonomy as especially detrimental to women," see Diana T. Meyers, *Self, Society, and Personal Choice* (New York: Columbia University Press, 1989).
8. Kant, *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck (Indianapolis, IN: Bobbs-Merrill Company, 1959); *The Doctrine of Virtue*, part II of the "Metaphysics of Morals," trans. Mary Gregor (Philadelphia: University of Pennsylvania Press, 1964), esp. p. 127.
9. Mill, *On Liberty*, in *Collected Works of John Stuart Mill*, vol. 18 (Toronto: University of Toronto Press, 1977), ch. I, III.
10. See, for example, Daniel Callahan, "Autonomy: A Moral Good, Not a Moral Obsession," *Hastings Center Report* 14 (October 1984): 40–42; Robert M. Veatch, "Autonomy's Temporary Triumph," *Hastings Center Report* 14 (October 1984): 38–40; and James F. Childress, "The Place of Autonomy in Bioethics," *Hastings Center Report* 20 (January/February 1990): 12–16.
11. See Barbara Herman, "Mutual Aid and Respect for Persons," *Ethics* 94 (July 1984): 577–602, esp. 600–602; Onora O'Neill, "Universal Laws and Ends-in-Themselves," *Monist* 72 (1989): 341–61.
12. See Daniel Callahan, "Autonomy: A Moral Good, Not a Moral Obsession"; and Colleen D. Clements and Roger C. Sider, "Medical Ethics' Assault Upon Medical Values," *Journal of the American Medical Association* 250 (Oct. 21, 1983): 2011–15.
13. See Bernard Lo et al., "Voluntary Screening for Human Immunodeficiency Virus (HIV) Infection: Weighing the Benefits and Harms," *Annals of Internal Medicine* 110 (May 1989): 727–33; and Martha S. Swartz, "AIDS Testing and Informed Consent," *Journal of Health Politics, Policy, and Law* 13 (Winter 1988): 607–21.
14. Childress, *Who Should Decide?* (New York: Oxford University Press, 1982), pp. 224–25. This case was prepared by Gail Povar, M.D.

15. See Bernard Gert and Charles Culver, "The Justification of Paternalism," *Ethics* 89 (January 1979): 199–210.
16. *Werth v. Taylor*, 190 Mich App 141 (1991).
17. *In re Estate of Dorone*, 502 A.2d 1271 (Pa. Super. 1985)
18. See Rebecca Dresser and John Robertson, "Quality of Life and Non-Treatment Decisions for Incompetent Patients: A Critique of the Orthodox Approach," *Law, Medicine, and Health Care* 17 (1989): 234–44.
19. See Allen E. Buchanan and Dan W. Brock, *Deciding for Others: The Ethics of Surrogate Decision Making* (Cambridge: Cambridge University Press, 1989), pp. 26–27. This book helped us correct some parts of our argument as presented in our third edition.
20. The analysis in this section has profited from discussions with Ruth R. Faden, Nancy M. P. King, and Dan Brock.
21. See the analysis of the core meaning in Charles M. Culver and Bernard Gert, *Philosophy in Medicine* (New York: Oxford University Press, 1982), pp. 123–26.
22. See *Lake v. Cameron*, 267 F. Supp. 155 (D.D.C. 1967).
23. *Pratt v. Davis*, 118 Ill. App. 161 (1905), aff'd, 224 Ill. 300, 79 N.E. 562 (1906).
24. See Daniel Wikler, "Paternalism and the Mildly Retarded," *Philosophy and Public Affairs* 8 (Summer 1979): 377–92.
25. A number of subtleties and needed qualifications in this analysis are discussed in an important paper by Kenneth F. Schaffner, "Competency: A Triaxial Concept," in *Competency*, ed. M. A. G. Cutter and E. E. Shelp (Dordrecht, the Netherlands: Kluwer Academic Publisher, 1991), pp. 253–81.
26. This case was prepared by P. Browning Hoffman, M.D., for presentation in the series of "Medicine and Society" conferences at the University of Virginia.
27. This schema is indebted to Paul S. Appelbaum, Charles W. Lidz, and Alan Meisel, *Informed Consent: Legal Theory and Clinical Practice* (New York: Oxford University Press, 1987), ch. 5; Ruth Macklin, "Some Problems in Gaining Informed Consent from Psychiatric Patients," *Emory Law Journal* 31 (Spring 1982): 345–74; Paul S. Appelbaum and Thomas Grisso, "Assessing Patients' Capacities to Consent to Treatment," *New England Journal of Medicine* 319 (December 22, 1988): 1635–38.
28. For additional ways in which values are incorporated, see Loretta M. Kopelman, "On the Evaluative Nature of Competency and Capacity Judgments," *International Journal of Law and Psychiatry* 13 (1990): 309–29. For conceptual and epistemic problems in all available tests, see E. Haavi Morreim, "Competence: At the Intersection of Law, Medicine, and Philosophy," in *Competency*, pp. 93–125, esp. 105–8.
29. See Willard Gaylin, "The Competence of Children: No Longer All or None," *Hastings Center Report* 12 (April 1982): 33–38, esp. 35; Buchanan and Brock, *Deciding for Others*, pp. 51–70; and Dan Brock, "Children's Competence for Health Care Decisionmaking," in *Children and Health Care*, ed. Loretta Kopelman and John Moskop (Boston: Kluwer Academic Publishers, 1989), pp. 181–212.
30. Buchanan and Brock, *Deciding for Others*, pp. 52–55. For elaboration and defense, see Brock, "Decisionmaking Competence and Risk," *Bioethics* 5 (1991): 105–12.
31. Related problems in the Buchanan-Brock analysis are discussed in Mark R. Wicclair, "Patient Decision-Making Capacity and Risk," *Bioethics* 5 (1991): 91–104, esp. p. 98 (and see p. 120 for an additional "Response").

32. Wendy Carlton, *"In Our Professional Opinion . . ." The Primacy of Clinical Judgment over Moral Choice* (Notre Dame, IN: University of Notre Dame Press, 1978), pp. 5–6.
33. See Jay Katz, *The Silent World of Doctor and Patient* (New York: The Free Press, 1984), pp. 86–87; and President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Making Health Care Decisions* (Washington, DC: U.S. Government Printing Office, 1982), vol. I, p. 15. Alan Weisbard, a proponent of this approach, suggests abandoning "informed consent" in favor of a different term, in "Informed Consent: The Law's Uneasy Compromise with Ethical Theory," *Nebraska Law Review* 65 (1986): 767.
34. See Charles W. Lidz et al., "Two Models of Implementing Informed Consent," *Archives of Internal Medicine* 148 (June 1988): 1385–89.
35. The analysis in this subsection is based in part on Faden and Beauchamp, *A History and Theory of Informed Consent*, ch. 8.
36. *Mohr v. Williams*, 95 Minn. 261, 265; 104 N.W. 12, 15 (1905).
37. This conclusion is vigorously defended by Weisbard, "Informed Consent: The Law's Uneasy Compromise with Ethical Theory," pp. 749–67.
38. Jay Katz, "Disclosure and Consent," in *Genetics and the Law II*, ed. A. Milunsky and G. Annas (New York: Plenum Press, 1980), pp. 122, 128; for revisions and amplifications, see Katz, "Physician-Patient Encounters 'On a Darkling Plain,' " *Western New England Law Review* 9 (1987): 207–26, and Alan Meisel, "A 'Dignitary Tort' as a Bridge between the Idea of Informed Consent and the Law of Informed Consent," *Law, Medicine, and Health Care* 16 (1988): 210–18.
39. See, for example, Alan Meisel and Loren Roth, "What We Do and Do Not Know about Informed Consent," *Journal of the American Medical Association* 246 (1981): 2473–77; President's Commission, *Making Health Care Decisions*, vol. II, pp. 317–410, esp. p. 318, and vol. I, ch. 1, esp. pp. 38–39; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report* (Washington, DC: DHEW Publication OS 78-0012, 1978), p. 10.
40. See, for example, *Planned Parenthood of Central Missouri v. Danforth*, 428 U.S. 52 at 67 n.8 (1976) (U.S. Supreme Court).
41. See the controversies catalogued in Hunter L. Prillaman, "A Physician's Duty to Inform of Newly Developed Therapy," *Journal of Contemporary Health Law and Policy* 6 (1990): 43–58.
42. *Moore v. Regents of the University of California*, 793 P.2d 479 (Cal. 1990) at 483.
43. See "Necessity and Sufficiency of Expert Evidence and Extent of Physician's Duty to Inform Patient of Risks of Proposed Treatment," *American Law Reports* 3d, 52 (1977): 1084; and "Physician's Duty to Inform of Risks," *American Law Reports* 3d, 88 (1986): 1010–25.
44. See *Largey v. Rothman*, 540 A.2d 504 (N.J. 1988), at 505.
45. See, for example, Charles Keown, Paul Slovic, and Sarah Lichtenstein, "Attitudes of Physicians, Pharmacists, and Laypersons Toward Seriousness and Need for Disclosure of Prescription Drug Side Effects," *Health Psychology* 3 (1984): 1–11, and Ruth R. Faden et al., "Disclosure of Information to Patients in Medical Care," *Medical Care* 19 (July 1981): 718–33.
46. See "Physician's Duty to Inform of Risks," pp. 1010–25 (and update, 50–60).
47. Ruth R. Faden and Tom L. Beauchamp, "Decision-Making and Informed Consent: A Study of the Impact of Disclosed Information," *Social Indicators Research* 7 (1980): 313–36.
48. Carl H. Fellner and John R. Marshall, "Kidney Donors—The Myth of Informed Consent," *American Journal of Psychiatry* 126 (1970): 1245–50, and "Twelve Kidney Donors," *Journal of the American Medical Association* 206 (1968): 2703–7.
49. See L. A. Siminoff and J. H. Fetting, "Factors Affecting Treatment Decisions for a Life-Threatening Illness: The Case of Medical Treatment of Breast Cancer," *Social Science and Medicine* 32 (1991): 813–18; David G. Scherer and N. D. Reppucci, "Adolescents' Capacities to Provide Voluntary Informed Consent," *Law and Human Behavior* 12 (1988): 123–41.
50. Gerald T. Roling et al., "An Appraisal of Patients' Reactions to 'Informed Consent' for Peroral Endoscopy," *Gastrointestinal Endoscopy* 24 (November 1977): 69–70.
51. L. A. Siminoff, J. H. Fetting, and M. D. Abeloff, "Doctor-Patient Communication about Breast Cancer Adjuvant Therapy," *Journal of Clinical Oncology* 7 (1989): 1192–1200.
52. The Oklahoma Supreme Court has been particularly vigorous in asserting this conception of the need for information. See *Scott v. Bradford*, 606 P.2d 554 (Okla. 1979) at 559 (together with *Masquat v. Maguire*, 638 P.2d 1105, Okla. 1981).
53. *Canterbury v. Spence*, 464 F.2d 772 (1977), at 785–89. See also *Wilson v. Scott*, 412 S.W.2d 299, 301 (Tex. 1967), and F. F. W. van Oosten, "The So-Called 'Therapeutic Privilege' or 'Contra-Indication': Its Nature and Role in Non-Disclosure Cases," *Medicine and Law* 10 (1991): 31–41.
54. *Thornburgh v. American College of Obstetricians*, 106 S.Ct. 2169, at 2199–2200 (1986) (White, J., dissenting).
55. Robert W. Allen, "Informed Consent: A Medical Decision," *Radiology* 119 (April 1976): 233–34.
56. See Kimberly A. Quaid et al., "Informed Consent for a Prescription Drug: Impact of Disclosed Information on Patient Understanding and Medical Outcomes," *Patient Education and Counselling* 15 (1990): 249–59.
57. James W. Lankton, Barron M. Batehelder, and Alan J. Ominsky, "Emotional Responses to Detailed Risk Disclosure for Anesthesia: A Prospective, Randomized Study," *Anesthesiology* 46 (April 1977): 294–96.
58. See Howard Brody, *Placebos and the Philosophy of Medicine: Clinical, Conceptual, and Ethical Issues* (Chicago: University of Chicago Press, 1980), pp. 10–11; and Herbert Benson and Mark Epstein, "The Placebo Effect: A Neglected Aspect in the Care of Patients," *Journal of the American Medical Association* 232 (1975): 1225.
59. Alan Leslie, "Ethics and Practice of Placebo Therapy," in *Ethics in Medicine*, ed. S. Reiser, A. Dyck, and W. Curran (Cambridge, MA: MIT Press, 1977), p. 242.
60. See L. C. Park et al., "Effects of Informed Consent on Research Patients and Study Results," *Journal of Nervous and Mental Disease* 145 (1967): 349–57.
61. Brody, *Placebos and the Philosophy of Medicine*, pp. 110, 113, *et passim*; Katz, *The Silent World*, pp. 189–95. For a defense of placebos, see Howard Spiro, *Doctors, Patients, and Placebos* (New Haven: Yale University Press, 1986).
62. Philip Levendusky and Loren Pankratz, "Self-Control Techniques as an Alternative to Pain Medication," *Journal of Abnormal Psychology* 84, no. 2 (1975): 165–68.

63. See Herbert C. Kelman, "Was Deception Justified—and Was It Necessary?" *Journal of Abnormal Psychology* 84 (1975): 172–74.
64. Sally E. McNagy and Ruth M. Parker, "High Prevalence of Recent Cocaine Use and the Unreliability of Patient Self-report in an Inner-city Walk-in Clinic," *Journal of the American Medical Association* 267 (February 26, 1992): 1106–8.
65. As Sissela Bok has argued in "Informed Consent in Tests of Patient Reliability," *Journal of the American Medical Association* 267 (February 26, 1992): 1118–19.
66. Stanley Milgram, *Obedience to Authority: An Experimental View* (New York: Harper & Row, 1974), pp. 3–4. See also Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67 (1963): 371–78; and "Some Conditions of Obedience and Disobedience to Authority," *Human Relations* 18 (1965): 57–76.
67. Milgram, "Subject Reaction: The Neglected Factor in the Ethics of Experimentation," *Hastings Center Report* 7 (October 1977): 19–21.
68. See Diana Baumrind, "Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience,' " *American Psychologist* 19 (1964): 421–23; Milgram, "Issues in the Study of Obedience: A Reply to Baumrind," *American Psychologist* 19 (1964): 848–52; and Steven C. Patten, "The Case That Milgram Makes," *Philosophical Review* 86 (July 1977): 350–64.
69. Milgram, "Subject Reaction," p. 19.
70. *Bang v. Charles T. Miller Hospital*, 251 Minn. 427, 88 N.W. 2d 186 (1958).
71. Franz J. Ingelfinger, "Informed (But Uneducated) Consent," *New England Journal of Medicine* 287 (August 31, 1972): 455–56. See also his "Arrogance," *New England Journal of Medicine* 303 (December 25, 1980): 1507–11.
72. See Paul R. Benson et al., "Information Disclosure, Subject Understanding, and Informed Consent in Psychiatric Research," *Law and Human Behavior* 12 (1988): 455–75.
73. The pioneering work was done by Amos Tversky and Daniel Kahneman. See "Choices, Values and Frames," *American Psychologist* 39 (1984): 341–50; "Judgment under Certainty: Heuristics and Biases," *Science* 185 (1974): 1124–31; "The Framing of Decisions and the Psychology of Choice," *Science* 211 (1981): 453–58.
74. Kahneman and Tversky, "Choices, Values and Frames," 344–46; and Tversky and Kahneman, "The Framing of Decisions."
75. S. E. Eraker and H. C. Sox, "Assessment of Patients' Preferences for Therapeutic Outcome," *Medical Decision Making* 1 (1981): 29–39; Barbara McNeil et al., "On the Elicitation of Preferences for Alternative Therapies," *New England Journal of Medicine* 306 (May 27, 1982): 1259–62.
76. See Jon F. Merz and Baruch Fischoff, "Informed Consent Does Not Mean Rational Consent," *The Journal of Legal Medicine* 11 (1990): 321–50; Baruch Fischoff, Paul Slovic, and Sarah Lichtenstein, "Knowing What You Want: Measuring Liable Values," in *Cognitive Processes in Choice and Decision Behavior*, ed. Thomas Wallsten (Hillsdale, NJ: Lawrence Erlbaum Associates, 1980), 117–41.
77. See Mark Siegler, "Critical Illness: The Limits of Autonomy," *Hastings Center* Report 12 (October 1977): 12–15, and the reply to Siegler by Jay Katz, *The Silent World*, pp. 156–59.
78. Ruth Faden and Alan Faden, "False Belief and the Refusal of Medical Treatment," *Journal of Medical Ethics* 3 (1977): 133–36.
79. *Cobbs v. Grant*, 502 P.2d 1, 12 (1972).

80. Baruch Brody, *Life and Death Decision Making* (New York: Oxford University Press, 1988), p. 22.
81. See Ralph J. Alfidi, "Controversy, Alternatives, and Decisions in Complying with the Legal Doctrine of Informed Consent," *Radiology* 114 (January 1975): 231–34.
82. William M. Strull, Bernard Lo, and Gerald Charles, "Do Patients Want to Participate in Medical Decisionmaking?" *Journal of the American Medical Association* 252 (1984): 2990–94.
83. See Joel Feinberg, *Social Philosophy* (Englewood Cliffs, NJ: Prentice-Hall, 1973), p. 48; *Harm to Self*, pp. 112–18.
84. Our formulation is indebted to Robert Nozick, "Coercion," in *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*, ed. Sidney Morgenbesser, Patrick Suppes, and Morton White (New York: St. Martin's Press, 1969), pp. 440–72, and Bernard Gert, "Coercion and Freedom," in *Coercion: Nomos XIV*, ed. J. Roland Pennock and John W. Chapman (Chicago: Aldine, Atherton Inc. 1972), pp. 36–37.
85. Paul S. Appelbaum and Loren H. Roth, "Treatment Refusal in Medical Hospitals," in President's Commission, *Making Health Care Decisions*, vol. II, p. 443; see also pp. 452, 462, 466.
86. See Charles W. Lidz and Alan Meisel, "Informed Consent and the Structure of Medical Care," in President's Commission, *Making Health Care Decisions*, vol. II, pp. 317–410.
87. Eliot Freidson, *The Profession of Medicine* (New York: Dodd, Mead & Co., 1970), p. 376.
88. See James H. Jones, *Bad Blood* (New York: The Free Press, 1981); David J. Rothman, "Were Tuskegee & Willowbrook 'Studies in Nature'?" *Hastings Center Report* 12 (April 1982): 5–7.
89. Henry K. Beecher, *Research and the Individual: Human Studies* (Boston: Little, Brown, and Co., 1970), p. 6.
90. See Charles W. Lidz et al., *Informed Consent: A Study of Decisionmaking in Psychiatry* (New York: Guilford Press, 1984), ch. 7, esp. pp. 110–11, 117–23.
91. The problems mentioned in this paragraph were examined in *Kaimowitz v. Department of Mental Health*, Civil No. 73-19434-AW (Circuit Court, Wayne County, Mich., July 10, 1973), at 31–32.
92. For the distinction between decisional autonomy and executional autonomy, see Bart J. Collopy, "Autonomy in Long Term Care," *The Gerontologist* 28, Supplementary Issue (June 1988): 10–17.
93. For this distinction between option right and mandatory right, see Joel Feinberg, "Voluntary Euthanasia and the Inalienable Right to Life," *Philosophy and Public Affairs* 7 (1978): 93–123.
94. For this case, see "If You Let Them, They'd Stay in Bed All Morning: The Tyranny of Regulation in Nursing Home Life," in *Everyday Ethics: Resolving Dilemmas in Nursing Home Life*, ed. R. A. Kane and A. L. Caplan (New York: Springer Publishing Co., 1990), ch. 7. For autonomy in long-term care, see other chapters in *Everyday Ethics* and a special journal issue devoted to this subject: *The Gerontologist* 28 (June 1988).
95. For a defense of negotiated consent, see Harry R. Moody, *Ethics in an Aging Society* (Baltimore: The Johns Hopkins University Press, 1992), ch. 8.
96. *Superintendent of Belchertown State School v. Saikewicz*, Mass. 370 N.E. 2d 417 (1977). For developments in related court opinions, see Sean M. Dunphy and John

H. Cross, "Medical Decision Making for Incompetent Persons: The Massachusetts Substituted Judgment Model," *Western New England Law Review* 9 (1987): 153–67.

97. *In re Quinlan*, 70 N.J. 10, 355 A.2d 647 (1976), at 663–64.
98. John A. Robertson, "Organ Donations by Incompetents and the Substituted Judgment Doctrine," *Columbia Law Review* 76 (1976): 65.
99. See George Annas, "Help from the Dead: The Cases of Brother Fox and John Storar," *Hastings Center Report* 11 (June 1981): 19–20. For an extremely charitable defense of courts' ascriptions of rights of self-determination to incompetents, see Alan Strudler, "Self-Determination, Incompetence, and Medical Jurisprudence," *The Journal of Medicine and Philosophy* 13 (1988): 349–65.
100. See the empirical study and analysis in Allison B. Seckler et al., "Substituted Judgment: How Accurate are Proxy Predictions?," *Annals of Internal Medicine* 115 (1991): 92–98.
101. In re *Conroy*, 486 A.2d 1209 (N.J. 1985). All quotations below are taken from this source.
102. See the court's explicit recognition and development of this point in re *Jobes*, 108 NJ 394, 529 A.2d 434 (1987).
103. In a series of decisions in 1987, the New Jersey Supreme Court extended its analysis to other types of cases (In re *Farrell*, 529 A.2d 404, In re *Jobes*, 529 A.2d 434, In re *Peters*, 108 N.J. 865).
104. *In the Matter of the Application of John Evans against Bellevue Hospital*, Supreme Court of the State of New York, Index No. 16536/87 (1987).
105. John Warren et al., "Informed Consent by Proxy: An Issue in Research with Elderly Patients," *New England Journal of Medicine* 315 (October 30, 1986): 1124–28, esp. 1127–28.
106. Steven H. Miles and Alison August, "Courts, Gender and 'The Right to Die,' " *Law, Medicine and Health Care* 18 (Spring-Summer 1990): 85–95.
107. Appelbaum, Lidz, and Meisel, *Informed Consent: Legal Theory and Clinical Practice*.
108. The classic case is *Strunk v. Strunk*, 445 S.W.2d 145 (Ky 1969), which considered these benefits in terms of a standard of substituted judgment.

4

Nonmaleficence

The principle of nonmaleficence asserts an obligation not to inflict harm intentionally. It has been closely associated in medical ethics with the maxim *Primum non nocere:* "Above all [or first] do no harm." This maxim is frequently invoked by health care professionals, yet its origins are obscure and its implications unclear. Often proclaimed the fundamental principle in the Hippocratic tradition of medical ethics, it is not found in the Hippocratic corpus, and a venerable statement sometimes confused with it—"at least, do no harm"—is a strained translation of a single Hippocratic passage.[1] Nonetheless, an obligation of nonmaleficence and an obligation of beneficence are both expressed in the Hippocratic oath: "I will use treatment to help the sick according to my ability and judgment, but I will never use it to injure or wrong them."

In this chapter we examine the principle of nonmaleficence and various attempts to specify its implications for biomedical ethics. In particular, we critically examine distinctions between killing and letting die, intending and foreseeing harmful outcomes, withholding and withdrawing life-sustaining treatments, and extraordinary and ordinary treatments. Many controversies in biomedical ethics surround the terminally ill and the seriously ill and injured. A framework is therefore needed for decisionmaking about life-sustaining procedures and assistance in dying. We defend a framework that would considerably alter current medical practice and guidelines for both competent and incompetent patients. At the center of the framework is an interpretation of the principle of nonmaleficence that sanctions rather than suppresses quality-of-life

judgments. This framework allows patients, guardians, and health care professionals under certain conditions to accept or refuse treatments after weighing the benefits and burdens of those treatments.

The Concept of Nonmaleficence

The Distinction between Nonmaleficence and Beneficence

A principle of nonmaleficence is recognized in many types of ethical theory, including utilitarian[2] and nonutilitarian writings.[3] Some philosophers join nonmaleficence and beneficence as a single principle. William Frankena, for example, treats the principle of beneficence as divisible into four general obligations, the first of which we will distinguish as the obligation of nonmaleficence and the other three of which we will refer to as obligations of beneficence:

1. One ought not to inflict evil or harm (what is bad).
2. One ought to prevent evil or harm.
3. One ought to remove evil or harm.
4. One ought to do or promote good.[4]

Frankena arranges these elements serially so that—other things being equal in a circumstance of conflict—the first takes moral precedence over the second, the second over the third, and the third over the fourth. Frankena acknowledges that the fourth element requires defense and qualification as a statement of *obligation*, for reasons we will address in Chapter 5.

If we try to encompass the ideas of benefiting others and not injuring them under a single principle, we will still be forced to distinguish, as Frankena does, among the various obligations embedded in this general principle. Although nonmaleficence and beneficence are similar and are often treated in moral philosophy as not sharply distinguishable, conflating them into one principle obscures relevant distinctions. Obligations not to harm others (for example, those prohibiting theft, disablement, and killing) are clearly distinct from obligations to help others (for example, by providing benefits, protecting interests, and promoting welfare). Obligations not to harm others are sometimes more stringent than obligations to help them, but obligations of beneficence are also sometimes more stringent than obligations of nonmaleficence. For example, the obligation not to injure others intuitively seems more stringent than the obligation to rescue them, but the obligation not to risk injury to research subjects through low-risk procedures is typically not as stringent as the obligation to rescue an injured research subject who underwent the procedures. If, in a particular case, the injury inflicted is very minor (swelling from a needlestick, say), but the benefit provided by rescue is major (a life-saving intervention, say), then the obligation of beneficence clearly takes priority over the obligation of nonmaleficence.

Many writers in ethics have maintained that one must accept substantial risks to one's safety in order not to cause harm to others, whereas acceptance of even moderate risks is not generally required to benefit others. But this claim too depends for its justification on particular situations, especially in professional ethics. For instance, public health officials in some countries cannot perform their jobs properly without undertaking at least moderate risks, such as exposing themselves to contagious diseases.

We might try to reformulate the idea of an increased stringency in nonmaleficence as follows: Generally, obligations of nonmaleficence are more stringent than obligations of beneficence; and, in some cases, nonmaleficence overrides beneficence when the best utilitarian outcome would be obtained by acting beneficently. For example, if a surgeon could save two innocent lives by killing a prisoner on death row to retrieve his heart and liver for transplantation, this outcome would have the highest net utility (in the circumstances), but it is not morally defensible. This formulation of the stringency of nonmaleficence has an initial ring of plausibility, especially if the act of benefiting involves committing a moral wrong. But again we should be cautious about axioms of priority. A utilitarian action does not necessarily take second place to an act of not causing harm. In cases of conflict, nonmaleficence is typically overriding, but the weights of these moral principles—like all moral principles—vary in different circumstances, and there thus can be no a priori rule that favors avoiding harm over providing benefit.

The claim that an order of priority exists among elements 1 through 4 in Frankena's scheme above is likewise difficult to sustain. Refraining from aiding another person (by not providing a good or by not preventing or removing harm) can be as morally wrong as inflicting a harm. Suppose the same harm occurs to X either by not assisting X, thereby permitting the harm to occur, or by inflicting the harm on X, and suppose the harm that is inflicted or allowed is equally intentional, equally sure to occur, and equally avoidable. Finally, suppose the actor is equally at negligible risk in the two scenarios. For example, the harm of death can be inflicted by killing a person with an injection or can be caused by failing to put a person on a respirator. The only difference in these cases is inflicting harm and refraining from assistance so as to allow harm, but this difference has no moral relevance. Thus, there is no moral difference between these two (or the above four) categories, and no order of priority among the categories.

It is preferable, we suggest, to distinguish the principles of nonmaleficence and beneficence conceptually in the following way, without proposing any normative ranking or hierarchical structure.

Nonmaleficence

1. One ought not to inflict evil or harm.

Beneficence

2. One ought to prevent evil or harm.
3. One ought to remove evil or harm.
4. One ought to do or promote good.

Each of these three forms of beneficence requires taking action by helping—preventing harm, removing harm, and promoting good—whereas nonmaleficence only requires intentionally refraining from actions that cause harm. Rules of nonmaleficence therefore take the form "Do not do X." Some philosophers accept only principles or rules that take a similarly proscriptive form. They interpret even rules of respect for autonomy as limited to rules of the form "Do not interfere with a person's autonomous choices." These philosophers reject all principles or rules that require helping, assisting, or rescuing other persons (although they recognize these norms as legitimate moral ideals). However, the mainstream of moral philosophy has not accepted such a *sharp* distinction between obligations of harming and helping—preferring instead to recognize and preserve the distinction in other ways. We will take this same path, and in Chapter 5 we will explain further the nature of the distinction and why conditions other than some form of *priority* for nonmaleficence can appropriately account for the distinction.

Legitimate disagreements have arisen about how to classify many actions under categories 1 through 4 as well as about the nature and stringency of the obligations that are involved in various circumstances. Consider, for example, the following case. Robert McFall was dying of aplastic anemia, and his physicians recommended a bone marrow transplant from a genetically compatible donor to increase his chances of living one additional year from twenty-five percent to a range of forty to sixty percent. The patient's cousin, David Shimp, agreed to undergo tests to determine his suitability to be a donor. After completing the test for tissue compatibility, he refused to undergo the test for genetic compatibility. He had changed his mind about donation. Robert McFall's lawyer asked a court to require that Mr. Shimp undergo the second test and donate his bone marrow if the test indicated a good match.[5]

Public discussion focused on whether David Shimp had an obligation of beneficence toward Robert McFall in the form of an obligation to prevent harm, to remove harm, or to promote McFall's welfare. McFall's lawyer contended (unsuccessfully) that even if Shimp did not have a legal obligation of beneficence to rescue his cousin, he had a legal obligation of nonmaleficence, which required that he not make McFall's situation worse. The lawyer argued that when Shimp agreed to undergo the first test and then backed out, he caused a

"delay of critical proportions" and violated the obligation of nonmaleficence. However, the judge ruled that Shimp did not violate any legal obligations but held that his actions were "morally indefensible."[6] This case illustrates the difficulties of identifying specific obligations under the principles of beneficence and nonmaleficence.

The Concept of Harm

The concept of nonmaleficence is frequently explicated using the terms *harm* and *injury*. *Injury* refers to harm, on the one hand, and to injustice, violation, or wrong, on the other.[7] The term *harm* has a similar ambiguity. "X harmed Y" might mean that X wronged Y or treated Y unjustly, or only that X thwarted, defeated, or set back Y's interests. Wronging involves violating someone's rights, whereas harming need not involve a violation. People are harmed without being wronged by diseases, acts of God, and bad luck; and people are wronged without being harmed whenever a wrongful action such as withholding promised information accidentally redounds to their benefit.[8] To explicate the principle of nonmaleficence, we will construe harm only in the second and normatively neutral sense of thwarting, defeating, or setting back the interests of one party by causes that include self-harming conditions as well as the (intentional or unintentional) actions of another party. Therefore, a *harmful* invasion by one party of another's interests may not be *wrong or unjustified*, although it is prima facie wrong. Some harmful actions are justifiable setbacks to another's interests—as, for example, in cases of justified criminal punishment and war, and even in cases of balancing competing interests. As with actions that punish other persons, rightness or wrongness depends on the strength of one's justification for the action. Perhaps even more importantly, what counts as a harm to one person may not be a harm at all to another person, because of their competing visions of what constitutes a setback to interests. We will return to this problem in the chapter on beneficence (pp. 289ff, 293ff, 301ff).

Some definitions of harm are so broad as to include setbacks to reputation, property, privacy, or liberty. Within this broad definition, trivial harms can be distinguished from serious harms by the order and magnitude of the interests affected. Other definitions with a narrower focus view harms exclusively as setbacks to physical and psychological interests, such as those in health and survival. Whether the broad or the narrow account is preferable is not critical for our discussion. We will concentrate on physical harms, including pain, disability, and death, without denying the importance of mental harms and setbacks to other interests. In particular, we will emphasize intending, causing, and permitting death or risk of death.

Rules Supported by the Principle of Nonmaleficence

Because of the many types of harm, the principle of nonmaleficence supports many more specific moral rules (although other principles will also occasionally be called on to help justify these rules). Typical examples of such rules include:[9]

1. "Do not kill."
2. "Do not cause pain or suffering to others."
3. "Do not incapacitate others."
4. "Do not cause offense to others."
5. "Do not deprive others of the goods of life."

Both the principle and its specifications in these moral rules are prima facie, not absolute. As noted above, some philosophers assign a priority in their system to principles and rules that prohibit infliction of harm, but we reject this ordering principle and all similar hierarchical arrangements.[10]

The Standard of Due Care

Obligations of nonmaleficence are obligations of not inflicting harms and not imposing risks of harm. A person can harm or place another person at risk without malicious or harmful intent, and the agent of harm may or may not be morally or legally responsible for the harms. In some cases agents are causally responsible for a harm when they do not intend or are even unaware of the harm caused. For example, if cancer rates are elevated at a chemical plant as the result of exposure to a chemical not previously suspected as a carcinogen, workers have been placed at risk by their employer, although the harm was not intentionally or knowingly caused.

In cases of risk imposition, law and morality recognize a standard of due care that specifies the principle of nonmaleficence. This standard can be met only if the goals sought justify the risks that must be imposed to achieve the goals. Grave risks require commensurately momentous goals for their justification, and emergencies justify risks that are not justified in nonemergency situations. For example, attempting to save lives after a major accident justifies the dangers created by speeding emergency vehicles. Negligence, a departure from the standard of due care toward others, includes intentionally imposing risks that are unreasonable as well as unintentionally but carelessly imposing risks. The term *negligence* applies to several forms of failure to meet obligations, including the failure to guard against risks of harm to others.[11] In treating negligence, we will concentrate on conduct that falls below a standard of due

care that is established by law or by morality to protect others from the careless or unreasonable imposition of risks.[12]

Courts often must determine responsibility and liability for harm because a patient, client, or customer seeks compensation for setbacks to interests or punishment of a responsible party, or both. Legal liability will not be considered here, but the legal model of responsibility for harmful action suggests a general framework that can be adapted to express the idea of moral responsibility for harm caused by health care professionals. The following are essential elements in a professional model of due care:

1. The professional must have a duty to the affected party.
2. The professional must breach that duty.
3. The affected party must experience a harm.
4. The harm must be caused by the breach of duty.[13]

Professional malpractice is an instance of negligence in which professional standards of care have not been followed.

For health care professionals, legal and moral standards of due care include proper training, skills, and diligence. In making services available, physicians accept the responsibility to observe these standards. If their conduct falls below these standards, they act negligently. If the therapeutic relationship proves to be harmful or unhelpful, malpractice occurs if and only if professional standards of care are not met. For example, in *Adkins v. Ropp*, the Supreme Court of Indiana considered a patient's claim that a physician had been negligent in removing foreign matter from the patient's eye and that, as a result, the eye became infected and blinded. The court held as follows:

> When a physician and surgeon assumes to treat and care for a patient, in the absence of a special agreement, he is held in law to have impliedly contracted that he possesses the reasonable and ordinary qualifications of his profession and that he will exercise at least reasonable skill, care and diligence in his treatment of him. This implied contract on the part of the physician does not include a promise to effect a cure and negligence cannot be imputed because a cure is not effected, but he does impliedly promise that he will use due diligence and ordinary skill in his treatment of the patient so that a cure may follow such care and skill, and this degree of care and skill is required of him, not only in performing an operation or administering first treatments, but he is held to the like degree of care and skill in the necessary subsequent treatments unless he is excused from further service by the patient himself, or the physician or surgeon upon due notice refuses to further treat the case.[14]

The customs, practices, and policies of the medical profession help establish applicable criteria of due care. For example, the Principles of Medical Ethics of the American Medical Association require physicians to provide "competent medical service" and to "continue to study, apply and advance scientific

knowledge.'' In some circumstances, ''a physician shall . . . obtain consultation, and use the talents of other health professionals when indicated.'' Although due care requirements cannot eliminate all mistakes or prevent all harms, they can reduce the probability of harmful outcomes in diagnosis and treatment.

The line between due care and care that falls below or exceeds what is due is often difficult to draw. Health risks can sometimes be reduced—in industry, say—by increased safety measures, by investing in epidemiological and toxicological studies, by educational or health promotional programs, by training programs, and the like. But a substantial question often remains about the lengths to which physicians, employers, and others must go to avoid or lower risks and thereby satisfy due care criteria. We will see below how this problem presents difficulties for determining the scope of obligations of nonmaleficence.

Traditional Distinctions and Rules Governing Nontreatment

Several guidelines have been developed in religious traditions, philosophical discourse, professional codes, and the law to specify requirements of nonmaleficence in health care, particularly with regard to treatment and nontreatment decisions. Some of these guidelines are thorough and helpful, but others should be either revised or replaced.

Several traditional guidelines draw heavily on the following distinctions:

1. Withholding and withdrawing life-sustaining treatment
2. Extraordinary (or heroic) and ordinary treatment
3. Artificial feeding and life-sustaining medical technologies
4. Intended effects and merely foreseen effects

We will argue that these distinctions are all untenable. They are distinctions without a relevant difference and should be replaced by the distinction between obligatory and optional means of treatment and by an account of the benefit–burden ratio. The venerable position that these traditional distinctions occupy in many professional codes, institutional policies, and writings in biomedical ethics does not supply an adequate reason for retaining them. Indeed, some of these distinctions are morally dangerous.

Withholding vs. Withdrawing Treatments

Much debate about the principle of nonmaleficence and forgoing life-sustaining treatments has centered on the omission–commission distinction, especially the distinction between withholding (not starting) and withdrawing (stopping) treatments. Many professionals and family members feel justified in withholding

treatments they never started, but not in withdrawing treatments already initiated. They sense that decisions to stop treatments are more momentous and consequential than decisions not to start them. Stopping a respirator, for example, seems to cause the person's death, whereas not starting the respirator seems like a prudent medical decision. But are these beliefs justifiable?

Consider the following case: An elderly man suffered from several major medical problems, including cancer, with no reasonable chance of recovery. Comatose and unable to communicate, he was being kept alive by antibiotics to fight infection and by an intravenous (IV) line to provide nutrition and hydration. No evidence indicated that he had expressed his wishes about life-sustaining treatments while competent, and he had no family members to serve as proxy decisionmakers. The staff quickly agreed on a ''no code'' or ''do not resuscitate'' (DNR) order, a signed order that cardiopulmonary resuscitation not be attempted if a cardiac or respiratory arrest occurred. In the event of a cardiac arrest, the patient would be allowed to die. The staff was comfortable with this decision because of the patient's overall condition and prognosis and because not resuscitating the patient could be viewed as withholding rather than withdrawing treatment.

Some members of the health care team thought that all medical treatments, including artificial nutrition, hydration, and antibiotics, should be stopped, because they were ''extraordinary'' or ''heroic.'' Others, perhaps a majority, thought it was wrong to stop these treatments once they had been started. A disagreement erupted about whether it would be permissible not to insert the IV line again if it became infiltrated—that is, if it broke through the blood vessel and began leaking fluid into surrounding tissue. Some who had opposed stopping treatment felt comfortable about not reinserting the IV line because they viewed the action as withholding rather than withdrawing. They emphatically opposed reinsertion if it required a cutdown (an incision to gain access to the deep large blood vessels) or a central line into the heart. Others viewed the provision of artificial nutrition and hydration as a single process and felt that inserting the IV line again was simply restarting or continuing what had been interrupted. For them, not restarting was equivalent to withdrawing, and thus (unlike withholding) morally wrong.[15]

Caregivers' discomfort about withdrawing life-sustaining treatments appears to reflect the view that such actions render them more responsible for, and therefore culpable for, a patient's death—whereas they are not responsible if they do not initiate a life-sustaining treatment. Another source of caregiver discomfort about withdrawing treatments is the conviction that starting a treatment often creates expectations that it will be continued, whereas stopping it appears to breach expectations, promises, or contractual obligations to the patient and family. Such wrong expectations and misleading promises can and should be avoided from the outset. The appropriate expectation or promise is

that caregivers will act in accordance with the patient's interests and wishes (as limited by defensible systems for the allocation of health care and defensible social rules about killing). Withdrawing a particular treatment, including life support, need not involve abandonment of the patient. It can follow the patient's directives and can be accompanied by other modes of care after a life-sustaining treatment is stopped.

Feelings of reluctance about withdrawing treatments are understandable, but the distinction between withdrawing and withholding treatments is morally untenable. The distinction is unclear, inasmuch as withdrawing can happen through an omission (withholding) such as not recharging batteries that power respirators or not putting the infusion into a feeding tube. In multistaged treatments, decisions not to start the next stage of a treatment plan can be tantamount to stopping treatment, even if the early phases of the treatment continue. Is such termination of the overall treatment withdrawing, withholding part of the treatment, or both?

Even if the distinction were clear, starting and stopping can both be justified, depending on the circumstances. Both starting and stopping can cause the death of a patient, and both can be instances of allowing to die. Courts recognize that a crime can be committed by omission if an obligation to act is present, just as a wrong can be committed by omission in medical practice. These judgments depend on whether a physician has an obligation to act in cases of either withholding or withdrawing treatment. In the *Spring* case (Case 5), the court raised a legal problem about continuing kidney dialysis as follows: "The question presented by . . . modern technology is, once undertaken, at what point does it cease to perform its intended function?" This court held that "a physician has no duty to continue treatment, once it has proven to be ineffective." The court emphasized the need to balance benefits and burdens to determine overall effectiveness.[16] Although legal responsibility cannot be equated with moral responsibility in such cases, this conclusion is consistent with the moral conclusions for which we are presently arguing.

Paradoxically, the moral burden of proof often should be heavier when the decision is to withhold than when it is to withdraw treatments.[17] In many cases, only after starting treatments will it be possible to make a proper diagnosis and prognosis as well as to balance prospective benefits and burdens. Uncertainty about outcomes can be reduced by such a trial period. Patients and surrogates often feel less stress and more in control if a decision to treat can be reversed or otherwise changed after it has been begun. Responsible health care, then, may require proposing a trial with periodic reevaluation. Caregivers then have time to judge the effectiveness of the treatment, and the patient or surrogate has time to evaluate its benefits and burdens. Not to propose or allow the test at all is morally worse than not trying.

The distinction between withholding and withdrawing treatment also can lead

to overtreatment in some cases—that is, to continuation of a treatment that is no longer beneficial or desirable for the patient. Less obviously, the distinction can lead to undertreatment. Patients and families worry about being trapped by biomedical technology that, once begun, cannot be stopped. To circumvent this problem, they become reluctant to authorize the technology, even when it could be beneficial. Health care professionals often exhibit the same reluctance. In one case, a seriously ill newborn died after several months of treatment, much of it against the parents' wishes, because a physician was unwilling to stop the respirator once it had been connected. Later it was reported that this physician was "less eager to attach babies to respirators now."[18] The distinction between withholding and withdrawing therefore can prevent patients from receiving medical benefits they should have.

We conclude that the distinction between withholding and withdrawing is morally irrelevant. Treatment can always permissibly be withdrawn if it can permissibly be withheld. This distinction, when combined with a reluctance to stop or a reluctance to start treatments under the assumption that they cannot then be stopped, creates dangerous situations for some patients. It also follows from the arguments in Chapter 3 that the patient has a right to forgo treatment at any time. Decisions about beginning or ending treatment should be based on considerations of the patient's rights and welfare, and, therefore, on the benefits and burdens of the treatment as judged by a patient or surrogate. Moreover, if a caregiver makes decisions about treatment using this irrelevant distinction, or allows a surrogate (without efforts at dissuasion) to make such a decision, the caregiver is morally blameworthy for any negative outcomes.

The felt importance of the distinction between not starting and stopping undoubtedly accounts for, although it does not justify, the ease with which hospitals and health care professionals have accepted "no code" or DNR orders. Hospital policies regarding cardiopulmonary resuscitation (CPR), a variety of interventions aimed at restoring function when a cardiac or respiratory arrest occurs, are particularly consequential, because cardiac arrest inevitably occurs in the dying process regardless of the underlying cause of death. CPR is commonly used in an attempt to prolong, at least briefly, the lives of patients who die in hospitals.

Policies regarding CPR are often independent of other policies about life-sustaining technologies, such as respirators, in part because many health care professionals view not providing CPR as withholding rather than withdrawing treatment. Their decisions to provide or not provide CPR are especially problematic when made without advance consultation with patients or their families.[19] DNR orders are often appropriate, and the option of such orders should be provided to patients or surrogates in a variety of circumstances, including terminal illness, irreversible loss of consciousness, and the likelihood of unmanageable cardiac or respiratory arrest. However, it is often unclear to hospi-

tal staffs, as well as to patients or their families, what, if anything, DNR orders imply about other levels of care and other technologies. For example, some patients with DNR orders still receive chemotherapy, surgery, and admission to the intensive care unit, whereas others do not. (Providing CPR where it is futile will be examined in Chapter 5.)

It is not justifiable to view decisions about CPR as different from decisions about other life-sustaining technologies. Neither the distinction between withholding and withdrawing treatments nor the distinction between ordinary and extraordinary means of treatment, as we will now argue, provides a justification.

Ordinary vs. Extraordinary Treatments

The distinction between ordinary and extraordinary treatments has been widely invoked both to justify and to condemn decisions to use or forgo life-sustaining treatments. The traditional rule is that extraordinary treatments can legitimately be forgone, but that ordinary treatments cannot legitimately be forgone. It has a prominent history in medical practice, judicial decisions, and Roman Catholic casuistry. This distinction has also been employed to determine whether an act that results in death counts as killing, especially as culpable killing. As developed by Roman Catholic theologians to deal with problems of surgery (prior to the development of antisepsis and anesthesia), this distinction has been used to determine whether a patient's refusal of treatment should be classified as suicide. Refusal of ordinary means of treatment was long considered suicide, but refusal of extraordinary means was not. Likewise, families and physicians did not commit homicide if they withheld or withdrew extraordinary means of treatment from patients.

Unfortunately, neither a long history nor contemporary precedent guarantees clarity or acceptability, and the distinction between ordinary and extraordinary means of treatment is both vague and morally unacceptable. Several problems surround the nature and purpose of the distinction. Throughout its history, several meanings and functions have been assigned to the distinction. *Ordinary* has often been taken to mean "usual" or "customary," whereas *extraordinary* has often been taken to mean "unusual" or "uncustomary." The ordinary has then been interpreted as the customary in medical practice, under either the professional practice standard discussed in Chapter 3 or the due care standard discussed earlier in this chapter. Treatments have been considered extraordinary if they are unusual or uncustomary for physicians to use in the relevant contexts. The terms thereby became attached to particular technologies.

The customary or usual in medical practice can be relevant to a moral judgment, but is not by itself sufficient or decisive. It is customary medical practice to treat a disease by a specific means, but whether this treatment should be repeated for a particular patient depends on the patient's wishes and condition as a whole, not alone on what is customary.[20] For example, treating pneumonia with antibiotics is usual, but it is morally optional for a patient who is irreversibly and imminently dying from cancer or AIDS. Ethical judgment is not reducible to professional custom, consensus, traditional codes, or oaths, as indispensable as these are for some professional contexts.

Criteria other than usual and unusual medical practice have also been proposed for the extraordinary procedures. These criteria include whether the treatment is simple or complex, natural or artificial, noninvasive or highly invasive, inexpensive or expensive, and routine or heroic. These substitutions are rarely analyzed with care and usually reduce to or mark no advance over *usual* and *unusual*. If a treatment is simple, natural, noninvasive, inexpensive, or routine, it is more likely to be viewed as ordinary (and thus obligatory) than if it is complex, artificial, invasive, expensive, or heroic (and thus optional). But these criteria are relevant only if some deeper moral considerations make them relevant. For instance, if a complex treatment is available and in accordance with the patient's wishes and interests, it is difficult to see why morally it should be distinguished from a simple treatment in accordance with the patient's wishes and interests. Similar confusions surround the criteria of the natural and the artificial. According to one study, physicians typically view respirators, dialyzers, and resuscitators as artificial; but they split evenly on intravenous feeding, and two-thirds view insulin, antibiotics, and chemotherapy as natural. They usually view mechanical systems as more artificial than drugs and other treatments.[21]

More consequential than these conceptual problems is whether such distinctions give sound moral guidance for treatment and nontreatment decisions. All treatments that fall into these classifications are sometimes beneficial for patients, sometimes burdensome; the principal consideration is whether a treatment is beneficial or burdensome, not its form. Thus, these distinctions appear irrelevant except insofar as they point to a quality-of-life criterion that requires balancing benefits against burdens. The need to balance benefits and burdens of treatments appears in the following influential exposition of the ordinary-extraordinary distinction:

> Ordinary means are all medicines, treatments, and operations which offer a reasonable hope of benefit and which can be obtained and used without excessive expense, pain, or other inconvenience. Extraordinary means are all medicines, treatments, and operations, which cannot be obtained or used without excessive expense, pain, or other inconvenience, or which, if used, would not offer a reasonable hope of benefit.[22]

If excessiveness is to be determined by the probability and magnitude of the benefit, as weighed against the likely burdens, this distinction will in the end reduce to the balance of burdens and benefits. If no reasonable hope of benefit

exists, then any expense, pain, or other inconvenience is excessive, and it is sometimes obligatory not to treat. If a reasonable hope of benefit exists, along with significant burdensomeness, the treatment is optional. Competent patients have a right to make decisions about treatment in light of their assessment of burdens and benefits, and for incompetent patients the treatment is not obligatory if a high level of burdensomeness is present. The ordinary–extraordinary distinction thus collapses into the balance between benefits and burdens, where the latter category includes immediate detriment, inconvenience, risk of harm, and other costs.

We conclude that the distinction between ordinary and extraordinary treatment is morally irrelevant and should be replaced by the distinction between optional and obligatory treatment, as determined by the balance of benefits and burdens to the patient.

Sustenance Technologies vs. Medical Technologies

In recent years there has been widespread debate about whether the distinction between *medical* technologies and *sustenance* technologies that supply nutrition and hydration using needles, tubes, catheters, and the like can legitimately be used to distinguish justified and unjustified forgoing of life-sustaining treatments. Some argue that technologies of caregiving for dispensing sustenance, such as artificially administered nutrition and hydration, are *nonmedical* means of maintaining life that are unlike optional forms of medical life-sustaining treatment, such as respirators and dialysis machines. To determine whether this distinction is more acceptable than the previous distinctions, we begin with three cases.

First, consider the case of a seventy-nine-year-old widow who had been a resident of a nursing home for several years. In the past she had experienced repeated transient ischemic attacks (caused by reductions or stoppages of blood flow to the brain). Because of progressive organic brain syndrome, she had lost most of her mental abilities and had become disoriented. She also had thrombophlebitis (inflammation of a vein associated with clotting) and congestive heart failure. Her daughter and grandchildren visited her frequently and loved her deeply. One day she suffered a massive stroke. She made no recovery, remaining obtunded and nonverbal, but she continued to manifest a withdrawal reaction to painful stimuli and exhibited some purposeful behaviors. She strongly resisted a nasogastric tube being placed in her stomach to introduce nutritional formulas and water. At each attempt she thrashed about violently and pushed the tube away. When the tube was finally placed, she managed to remove it. After several days on intravenous lines, the sites for inserting IV lines were exhausted. The staff debated whether to take further "extraordinary" measures to maintain fluid and nutritional intake for this elderly patient who had failed to improve and was largely unaware and unresponsive. After lengthy discussions with nurses on the floor and with the patient's family, the physicians in charge reached the conclusion that they should not provide further IVs, cutdowns, or a feeding tube. The patient had minimal oral intake and died quietly the following week.[23]

Second, in a ground-breaking case in 1976 the New Jersey Supreme Court held that it was permissible for a guardian to disconnect Karen Ann Quinlan's respirator and allow her to die.[24] After the respirator was removed, she lived for almost ten years, protected by antibiotics and sustained by nutrition and hydration provided through a nasogastric tube. Unable to communicate, she lay comatose in a fetal position, with increasing respiratory problems, bedsores, and weight loss (from 115 to 70 pounds). A moral issue developed over the course of those ten years. If it is permissible to remove the respirator, is it permissible for the same reasons to remove the feeding tube? Several Roman Catholic moral theologians advised the parents that they were not morally required to continue medically administered nutrition and hydration (MN&H) or antibiotics to fight infections. Nevertheless, the Quinlans continued MN&H because they believed that the feeding tube did not cause pain, whereas the respirator had.

Third, while Karen Quinlan lingered, the same state supreme court faced another case involving artificial nutrition and hydration in which a guardian requested withdrawal of MN&H for an eighty-four-year-old nursing home resident. The court held that the provision of nutrition and hydration through nasogastric tubes and other medical means is not always legally required.[25] A similar decision was soon reached in the Brophy case in Massachusetts, involving a forty-nine-year-old man who had been in a persistent vegetative state for more than three years.[26] Courts have since increasingly maintained that no relevant difference distinguishes MN&H from other life-support measures. They have viewed MN&H as a medical procedure subject to the same standards as other medical procedures and thus sometimes unjustifiably burdensome.[27] A similar discussion about MN&H has surfaced in treatment decisions about newborns who are seriously ill or severely disabled.

Whether one is confronted with an elderly woman, a twenty-year-old youth, or an infant, the same moral question is present: Should such medical procedures be construed as obligatory or as optional, and under which circumstances?[28] We maintain that MN&H may justifiably be forgone in some circumstances, as is true of other life-sustaining technologies. The chief premises in our argument are (1) that no morally relevant difference exists between various life-sustaining technologies and (2) that the right to refuse medical treatment is not contingent on the type of treatment. We find no reason to believe that MN&H is always an essential part of palliative care or that it necessarily constitutes, on balance, a beneficial medical treatment.

Although our view is consistent with many recent court decisions, professional codes, and philosophical arguments, it remains controversial. Philosopher G. E. M. Anscombe contends that, "For wilful starvation there can be no excuse. The same can't be said quite without qualification about failing to operate or to adopt some course of treatment."[29] C. Everett Koop, former Surgeon General of the United States, denounces practices of allowing newborns to die (including by omitting MN&H) as infanticide by "starving a child to death;"[30] he likewise condemns similar practices for adults as intentional acts of killing that amount to active euthanasia because they cause a preventable death.[31] Others maintain that MN&H are not relevantly similar to other treatments in medicine and that no moral principle requires or permits the judgments about MN&H that are legitimately made about other medical procedures. From this perspective, although it is legitimate to omit some forms of treatment, MN&H cannot justifiably be omitted.

Three main arguments have been advanced in defense of this position, and each directly challenges our position. The first argument holds that MN&H are required because they are necessary for the patient's comfort and dignity. This view underlies the controversial rule once proposed by the U.S. Department of Health and Human Services for treatment of disabled newborns: "The basic provision of nourishment, fluids, and routine nursing care is a fundamental matter of human dignity, not an option for medical judgment."[32] This rule includes all means of providing nutrition and hydration. A similar conviction about patient comfort and dignity underlies the exclusionary clause in several natural death acts that disallow advance directives about artificial nutrition and hydration, while permitting advance directives for nonsustenance procedures that prolong life.[33]

A second argument that MN&H are never optional focuses on symbolic significance. Medical professionals generally find it intuitively devastating to "starve" someone. Provision of nutrition and hydration symbolizes the essence of care and compassion in medical as well as nonmedical contexts. As Daniel Callahan puts it, feeding the hungry and nursing by nourishing are "the rudimentary healing gesture" and "the perfect symbol of the fact that human life is inescapably social and communal."[34] Our experiences of thirst and hunger enhance this symbol. Thirst and hunger are uncomfortable, and we infer that severe malnutrition and dehydration must involve extreme agony.

The third argument is a version of the wedge or slippery slope argument considered later in this chapter. The controlling idea is that policies of not providing MN&H will lead to adverse consequences because society will not be able to limit decisions about MN&H to legitimate cases, especially under pressures for cost containment in health care. Whereas "death with dignity" first emerged as a compassionate response to the threat of overtreatment, patients now face the threat of undertreatment because of pressures to contain the

escalating costs of health care. Such concerns about psychological and social barriers focus on a slide from acting in the patient's interests to acting in the society's interests, from considering the patient's quality of life to considering the patient's value for society, from decisions about terminally ill patients to decisions about nondying patients, from letting die to killing, and from cessation of artificial feeding to cessation of natural feeding. The fear is that the "right to die" will be transformed into the "obligation to die," perhaps against the patient's wishes and interests.[35]

We have several reservations about these three arguments. Whether the rationale is life prolongation or patient comfort and dignity, an absolute requirement to provide medically assisted nutrition and hydration has serious drawbacks. Procedures of MN&H themselves sometimes involve risks of harm, discomfort, and indignity, such as pain from a central IV and physical restraints that prevent patients from removing the lines or tubes. Evidence also indicates that patients who are allowed to die without artificial hydration sometimes die more comfortably than patients who receive such hydration. It is often misleading to project the common experience of hunger and thirst on a dying patient who is malnourished and dehydrated. Malnutrition is not identical with hunger; dehydration is not identical with thirst; and starvation is very different from acute dehydration in a medical setting. Feelings of hunger, thirst, dryness of the mouth, and related problems also can be alleviated by other means, such as ice on the lips, without introducing MN&H.[36]

For some patients the burdens of MN&H outweigh their benefits, and no one should deprive them of the right to refuse treatment. The obligation to care for patients entails provision of treatments that are in accordance with their preferences and interests (within the limits set by just allocation policies), not the provision of treatments because of what they symbolize in the larger society. In an approach that represents a compromise for physicians who want to engage in symbolically significant actions, while also acting in accord with the patient's wishes and interests, some physicians start and continue IV lines at a rate that will result in dehydration over time.[37] This approach is both risky and deceptive. These physicians fail to acknowledge their final objective, which is that the patient will become dehydrated and malnourished, and die as a result. The act resulting in death is intentional, and death at the time it occurs is a foreseen and avoidable consequence.

The fears underlying the third argument about the slippery slope are legitimate and troubling because of uncertainties about whether lines can be drawn and maintained in order to prevent abuses. Perhaps eighty percent of the over two million people who die each year in the United States die in nursing homes or hospitals under the care of strangers, often at considerable cost to their families and to the society.[38] These and other patients in long-term care are vulnerable, and we should be concerned about the potential loss of broad moral com-

mitments that form the cement of our social universe. However, no evidence exists that the protection of these patients requires that MN&H be provided in all circumstances or that the emotions underlying the symbol of providing nutrition and hydration are either necessary or sufficient to avert social disaster.

We conclude that it is sometimes legitimate to remove MN&H, paving the way for death and that the presumption in favor of MN&H can be successfully rebutted under the following conditions: (1) The procedures are highly unlikely to improve nutritional and fluid levels. (2) The procedures will improve nutritional and fluid levels, but the patient will not benefit (for example, in cases of anencephaly or permanent vegetative state). (3) The procedures will improve nutritional and fluid levels and the patient will benefit, but the burdens of MN&H outweigh the benefits. For example, for a severely demented patient, essential physical restraints can cause fear and discomfort, especially as the patient struggles to break free of the restraints.

Intended Effects vs. Merely Foreseen Effects

We have thus far rejected or revised several traditional rules and distinctions in medical ethics. Another venerable attempt to specify the principle of nonmaleficence appears in the rule of double effect (RDE), often called the principle or doctrine of double effect. This rule too has a pivotal distinction, viz. between intended effects (or consequences) and foreseen effects (or consequences). The RDE is invoked to justify claims that a single act having two foreseen effects, one good and one harmful (such as death), is not always morally prohibited if the harmful effect is not intended.[39]

Functions and conditions of the RDE. The RDE finds its natural setting in ethical theories that view certain actions as intrinsically and absolutely wrong, including theories that prohibit the direct infliction of harm on innocent persons.[40] The RDE is one attempt to specify the conditions of the principle of nonmaleficence for situations in which an agent cannot avoid all harms and still achieve important goods.

As an example of the use of the RDE, consider a patient experiencing terrible pain and suffering who asks a physician for help in ending his life. If the physician directly kills the patient to end the patient's pain and suffering, the patient's death is caused intentionally as a means to relieve pain and suffering. But suppose that the physician could provide medication to relieve the patient's pain and suffering at a substantial risk that the patient will die earlier as a result of the medication. If the physician refuses to administer the toxic analgesia, the patient will be harmed by continuing pain and suffering; if the physician provides the medication, the patient's death may be hastened. Under the RDE, the physician's provision of medication must be intended to relieve grave pain

and suffering and must not be intended to hasten death. If no intention of a lethal effect exists, this act is not prohibited by the principle "Do not directly and intentionally harm the innocent."

According to classical formulations of the RDE, four conditions or elements must be satisfied for an act with a double effect to be justified. Each is a necessary condition, and together they form sufficient conditions of morally permissible action:[41]

1. *The nature of the act.* The act must be good, or at least morally neutral (independent of its consequences).
2. *The agent's intention.* The agent intends only the good effect. The bad effect can be foreseen, tolerated, and permitted, but it must not be intended.
3. *The distinction between means and effects.* The bad effect must not be a means to the good effect. If the good effect were the direct causal result of the bad effect, the agent would intend the bad effect in pursuit of the good effect.
4. *Proportionality between the good effect and the bad effect.* The good effect must outweigh the bad effect. The bad effect is permissible only if a proportionate reason is present that compensates for permitting the foreseen bad effect.

Controversy has surrounded all four of these traditional conditions. We begin our analysis by considering four cases of what many call therapeutic abortion (limited to protection of maternal life in these examples). (A) A pregnant woman has cancer of the cervix; a hysterectomy is needed to save her life, but it will result in the death of the fetus. (B) A pregnant woman has an ectopic pregnancy—the nonviable fetus is in the fallopian tube—and removal of the tube, which will result in the death of the fetus, is medically indicated to prevent hemorrhage. (C) A pregnant woman has a serious heart disease that will probably result in her death if she attempts to carry the pregnancy to term. (D) A pregnant woman in difficult labor will die unless a craniotomy (crushing the head of the unborn fetus) is performed. Official Roman Catholic teaching and many moral philosophers and theologians hold that actions that produce fetal deaths in cases A and B sometimes satisfy the four conditions of the RDE and therefore are morally acceptable, whereas the actions that produce fetal deaths in cases C and D never meet the conditions of the RDE and therefore are morally unacceptable.[42]

In the first two cases, according to the RDE, a physician undertakes a legitimate medical procedure aimed at saving the pregnant woman's life with the foreseen, but unintended, result of fetal death. Viewed as side effects that are not intended (rather than as ends or means), these fetal deaths can be justified by a proportionately grave reason (saving the pregnant woman). In both cases

C and D, the action of killing the fetus as a means to save the pregnant woman's life requires intending the death (even if the death is not desired). Therefore, in those cases it is not permissible to consider proportionality.

Critics of the RDE contend that it is difficult and perhaps impossible to establish a morally relevant difference between cases such as A (hysterectomy) and D (craniotomy) through RDE conditions. In neither case does the agent want or desire the death of the fetus, and the descriptions of the acts in these cases do not indicate morally relevant differences. It is not clear why craniotomy is killing the fetus rather than crushing the skull of the fetus with the unintended result that the fetus dies. It is also not clear why in the hysterectomy case the death is foreseen but not intended. A proponent of the RDE must have a practicable way to distinguish the intended from the merely foreseen, but it has proved difficult to draw defensible moral lines between various cases like the hysterectomy and craniotomy cases. Some modern reformulations of the RDE (especially those emphasizing the fourth condition) even permit craniotomies to save the pregnant woman's life because of the proportional value of the woman's life.[43]

Critique of the RDE. Adherents of the RDE need an account of intentional actions and intended effects of action (intentionally causing or allowing) that properly distinguishes them from nonintentional actions and unintended effects (foreseeably causing or allowing). The literature on intentional action is itself highly controversial and focuses on diverse conditions such as volition, deliberateness, willing, reasoning, and planning. One of the few widely shared views in this literature is that intentional actions require an agent's plan—a blueprint, map, or representation of the means and ends proposed for the execution of an action.[44] For an action to be intentional, it must correspond to the agent's conception of how it was planned to be performed.

Alvin Goldman uses the following example in an attempt to prove that merely foreseen effects are unintentional.[45] Imagine that Mr. G is taking a driver's test to prove competence. He comes to an intersection that requires a right turn and extends his arm to signal for a turn, although he knows it is raining and his hand will become wet. According to Goldman, Mr. G's signaling for a turn is an intentional act. By contrast, his getting a wet hand is an unintended effect or "incidental by-product." The defender of the RDE elects a similarly narrow conception of what is intended in order to avoid the unacceptable conclusion that an agent intentionally brings about all the consequences of an action that the agent foresees. The defender distinguishes between acts and effects, and then between (1) effects that are desired or wanted and (2) effects that are foreseen but not desired or wanted. The latter effects are viewed in RDE as foreseen, but not intended.

However, it is better to discard the language of desiring and wanting altogether, and to say that these effects are "tolerated."[46] These effects are not so undesirable that the actor would choose not to perform the act that results in them, and they are a part of the plan of an intentional action. If we use a model of intentionality based on what is *willed* rather than what is *wanted*, intentional actions and intentional effects include any action and any effect willed in accordance with a plan, including tolerated as well as wanted effects.[47] On this conception, a physician can desire not to do what he intends to do, in the same way that we can be willing to do something but, at the same time, be reluctant to do it or even detest doing it. Undesirable effects or risks of harm that attend particular procedures usually fall into this category. Under this conception of intentional acts and intended effects, the distinction between what is intended and what is merely foreseen is not viable.[48]

Thus, a person who knowingly and voluntarily acts to bring about an effect brings about that effect intentionally. The effect is intended, although the person did not desire it, did not will it for its own sake, or did not intend it as the goal of the action. For example, if a man enters a room and flips a switch that he knows turns on both a light and a fan, but desires only to activate the light, he cannot say that he activates the fan unintentionally. Although the fan makes an obnoxious whirring that he desires to avoid, it would be conceptually mistaken to say that he unintentionally brought about the obnoxious sound by flipping the switch.

Finally, the moral relevance of the RDE and its distinctions must be considered. Is it plausible to distinguish morally between intentionally causing the death of a fetus by craniotomy and intentionally removing a cancerous uterus that causes the death of a fetus? In both actions, the intention is to save the woman's life with knowledge that the fetus will be lost. No agent desires a bad result (the fetus's death) for its own sake, and none would have tolerated the bad result if its avoidance were morally preferable. Each party accepts the bad effect only because it cannot be eliminated without losing the good effect. Accordingly, the agents in our various examples above do not appear to want, will, or intend in ways that make a moral difference.

In the standard interpretation of the RDE, the fetus's death is a *means* to saving a woman's life in the unacceptable case, but merely a *side effect* in the acceptable case; a means is intended, whereas a side effect need not be. But this approach seems to allow almost anything to be foreseen as a side effect rather than intended as a means (although this is not to claim that we can create or direct intentions as we please). For example, in the craniotomy case, the surgeon might not intend the death of the fetus but only intend to remove it from the birth canal. The fetus will die, but is this outcome more than an unwanted and (in double effect theory) unintended consequence?[49]

There may be a way out of these puzzles for defenders of the RDE, but it is doubtful that the way has yet been found.[50] Meanwhile, other criticisms need

to be answered. Critics who accept a broadly consequentialist approach argue that unacceptable consequences must be deemed acceptable in the doctrine of double effect. For example, slowly causing death by administering medication to reduce pain can involve painful days or weeks of a life that a patient wishes to end, whereas a more active means to death, such as a larger and lethal dose of the same painkiller, would end life more quickly and with less pain. Critics charge that the RDE forces its exponents to accept less humane methods of ending human life than should be provided.

One constructive effort to retain an emphasis on intention without entirely abandoning or neglecting the point of the RDE focuses on the way actions display a person's motives and character. From this perspective, the core issue is whether a person's conduct flows from a proper motivational structure and a good character. Often in evaluating persons we are more concerned with their *motivation* to perform an action (why they performed the action) than with their *intention* in performing the action (what they planned to do). The intention to kill another person may be less relevant morally than the motive in doing so—for example, self-defense or defense of an innocent third party. When juries hear cases of mercy killing, the intention to kill is morally and legally pertinent, but the crucial moral assessment is often the motive behind the killing, not the intention to kill. The motive to relieve suffering is, of course, entirely consistent with the intention to kill.

In the case of performing a craniotomy to save a pregnant woman's life, the action need not be motivated by a disregard for human life or by a positive desire to end it. A physician may not *want* or *desire* the death of the fetus in this case and may regret performing a craniotomy, just as much as in the case of removing a cancerous uterus. Such facts about the physician's motivation and character can make a decisive difference to a moral assessment of the action and the agent. But the RDE is unable to reach this conclusion on its own. In effect, our proposal to focus on motivation transforms the RDE into another moral framework. We develop the acceptable features of this framework in Chapter 8.

Even if one accepts the RDE, it will be irrelevant in many pressing problems about harm and killing that are currently under discussion in biomedical ethics, including the issues surrounding assisted suicide and euthanasia that we consider later in this chapter. The RDE is fashioned exclusively for cases with both a bad and a good effect, but often the central matter in dispute is whether an effect such as death is bad. Nothing in the RDE decides this issue. For example, one cannot decide from the RDE whether voluntary active euthanasia produces a bad effect or a good effect, because this premise must be defended or rejected on independent grounds.

Some parts of the RDE are perfectly acceptable—for example, the rule that a harmful effect is justifiably allowed only if a proportionately weighty good

will probably be brought about. But as we will now see, this latter rule can be put to many uses in biomedical ethics beyond those permitted by the full RDE.

Optional Treatments and Obligatory Treatments

We have now rejected several of the leading distinctions and rules about forgoing life-sustaining treatment sanctioned by various traditions of medical ethics. We do not deny that some of these traditional distinctions are deeply embedded in the beliefs and guidelines of many health care professions, as well as in the policies and practices of many health care institutions. However, the key question is whether the health care professions and institutions need to revise their traditional beliefs to accommodate a broader and more compelling moral perspective, including patients' rights of autonomy.

We will now propose a replacement distinction—viz. between obligatory and optional treatments—and explain why this distinction is more suitable. In doing so, we will rely heavily on an analysis of quality of life that is largely incompatible with the distinctions we have rejected. The following categories are central to our arguments:

I. Obligatory to Treat (Wrong Not to Treat)
II. Optional Whether to Treat
  A. Neutral (Neither Required nor Prohibited)
  B. Supererogatory (Surpassing Obligation)
III. Obligatory Not to Treat (Wrong to Treat)

Most ethical discussions have focused on I and II, with only scant attention to different interpretations of II. A treatment is optional, under IIA, if it is morally neutral whether a physician provides it, a surrogate authorizes or refuses it, and the like. It is optional, under II B, if providing it would be supererogatory and therefore praiseworthy, while not providing it would not be morally blameworthy. These terms are used to indicate a variety of actions, including those involving expenditure of additional time, effort, energy, and resources when a very limited chance of success is present. (For a discussion of the terms *supererogatory* and *praiseworthy*, see Chapter 8.)

Under category III, is it ever wrong to treat (or obligatory not to treat)? The principles of nonmaleficence and beneficence establish a presumption in favor of providing life-sustaining treatments for sick and injured patients, but they also indicate conditions for rebutting that presumption. In addition, life-sustaining treatments sometimes violate patients' interests. For example, pain can be so severe and physical restraints so burdensome that they outweigh anticipated benefits such as brief prolongation of life. In these circumstances, providing the treatment is sometimes inhumane or cruel and therefore in viola-

tion of the principle of nonmaleficence. It will often—perhaps usually—be difficult to determine the balance of benefits and burdens to the incompetent patient, particularly when he or she has never lived as a competent person expressing values. However, the burdens can so outweigh the benefits to the incompetent patient that the treatment is wrong rather than optional, just as in the case of a competent patient who refuses treatment.

We reserve a systematic treatment of cost–benefit and risk–benefit assessment for the next chapter. Our concern at present is with substantive standards that distinguish obligatory and optional treatments. Competent patients who can make informed and voluntary choices should have more latitude than other parties in balancing benefits and burdens and in accepting and refusing treatment. As noted throughout this chapter, the incompetent patient's vulnerability to harm sometimes requires actions, based on principles of nonmaleficence and beneficence, that would violate respect for a competent patient's autonomy unless authorized by that patient.

Conditions for Overriding the Prima Facie Obligation to Treat

Several conditions justify decisions to omit treatment by patients, surrogates, or health care professionals. We introduce these conditions in this section.

Futile or pointless treatment. Treatment is not obligatory when it offers no benefit to the patient because it is pointless or futile. Several treatments fit this description. For example, if a patient is dead, although still on a respirator, he or she can no longer be harmed by cessation of treatment, and a standard of medical best interests does not dictate treatment. However, in some religious and personal belief systems, a patient is not considered dead according to the criteria recognized in health care institutions. For example, if heart and lung function can be maintained, some religious traditions hold that the person is not dead, and the treatment is therefore not futile, even if it is deemed futile by health care professionals. This is the tip of an iceberg of controversies that surround the notion of futility.

Typically we think of the term *futile* as referring to a situation in which patients who are irreversibly dying have reached a point at which further treatment provides no physiological benefit or is hopeless and becomes optional, although palliative interventions (those intended to alleviate discomfort, pain, and suffering, but not to cure) may need to be continued. This model, however, covers only a narrow range of cases that have been *labeled* futile in the literature on the subject. All of the following have been referred to as futile treatments: whatever is highly unlikely to be efficacious (statistically the odds of success are exceedingly small), a low-grade outcome that is virtually certain (qualitatively the results are expected to be exceedingly poor), whatever is highly likely to be more burdensome than beneficial, and whatever is completely speculative because it is an untried "treatment." Thus, the term *futility* is now used to cover both situations of predicted impossibility and situations in which there are competing interpretations of probabilities and competing value judgments such as a balance of probable benefits and burdens (as we discuss below and in Chapter 5).[51] This situation of equivocation and ambiguity suggests that the term *futility* generally should be avoided in favor of more precise language.

Ideally in cases involving either those who are dead or those who are irreversibly dying, objective medical factors and expert judgment are central. Realistically, though, this ideal is difficult to satisfy in setting criteria of futility and in making judgments of futility. Disagreement often exists in the medical community, and conflicts may arise from a family's belief in a possible miracle, a religious tradition's insistence on doing everything in such a circumstance, and so forth. It is sometimes difficult to know whether a judgment of futility is based on a probabilistic prediction of failure or on something closer to medical certainty. If an elderly patient has a one percent chance of surviving an arduous and painful regimen, one physician may call the procedure futile and another may view survival as an unlikely outcome but still a possibility that should be considered. We here encounter a value judgment about what is worth the effort, as well as a judgment based on scientific knowledge. "Futility" is typically used to express a combined value judgment and scientific judgment, although many people mistakenly construe judgments of futility as value free.

Writings in biomedical ethics that discuss futility often focus on the patient's or surrogate's right to refuse futile treatment. However, circumstances have increasingly appeared in which the question is whether the physician may or should refuse to provide some treatment. The fact that a treatment is futile is often said to change the physician's moral relationship to patients or surrogates. The physician is not required to provide such treatment and sometimes is not required to discuss the treatment. These circumstances commonly involve incompetent persons, especially patients in a persistent vegetative state (PVS), where physicians or hospital policies impose decisions to forgo life-support on patients or surrogates. Increasingly hospitals are adopting policies explicitly aimed at denying therapies that are judged futile by physicians, especially after the therapy has been tried for a reasonable period of time.

The possibility of judgmental error by physicians should lead to caution in formulating these policies, but, at the same time, unreasonable demands by patients and families should not preclude reasonable policies by health care institutions. Here, as well as elsewhere, respect for the autonomy of patients is not a trump that allows them alone to determine whether a treatment is required or is futile. In one case, Mr. C. was irreversibly dying from emphysema and

insisted on having his life prolonged as long as possible by all available means. He demanded aggressive treatment, although the staff considered the treatment futile. When he became unconscious, his family and the staff had to decide whether to respect their earlier agreement with him or let him die. Without a prior statement of Mr. C.'s wishes, there would be no moral difficulty in terminating treatment that was only prolonging his dying. But even with the prior agreement, following Mr. C.'s previous wishes might not be justified because of the combination of futility and limited health care resources.[52] For instance, if other patients who were not irreversibly dying could not otherwise gain access to the ventilator and space in the intensive care unit, we would not be obligated to continue treatment.

The upshot is that a pointless or futile treatment, in the sense of a treatment that has no chance of being efficacious, is morally optional but that other putatively futile treatments are often not optional.

Burdens of treatment outweigh benefits. A mistaken assumption about law and ethics sometimes found in medical codes is that life-sustaining treatments may be terminated only if a patient is terminally ill. If the patient is not terminally ill, life-sustaining medical treatment is still not obligatory if its burdens outweigh its benefits to the patient. Medical treatment for those not terminally ill is sometimes optional, although it could prolong life for an indefinite period and the patient is incompetent and has left no advance directive.[53]

The principle of nonmaleficence does not imply the maintenance of biological life, nor does it require the initiation or continuation of treatment without regard to the patient's pain, suffering, and discomfort. In Case 5, seventy-eight-year-old Earle Spring developed numerous medical problems, including chronic organic brain syndrome and kidney failure. The latter problem was controlled by hemodialysis. Although several aspects of this case are in dispute—such as whether Spring was conscious, aware, and able to express his wishes—there is at least a plausible argument that the family and health care professionals were not morally obligated to continue hemodialysis, because of the balance of benefits and burdens to the patient. This case, like many others, is complicated by the fact that the family had a conflict of interest because of their obligations both to pay mounting and burdensome health care costs and to make judgments about the patient's best interests.

Few decisions are more momentous than those to withhold or withdraw a medical procedure that sustains life. But in some cases it is unjustified for surrogates and clinicians to begin or to continue therapy knowing that it will produce a greater balance of pain and suffering for a patient incapable of choosing for or against such therapy. As the supreme judicial court of Massachusetts once held, "the 'best interests' of an incompetent person are not necessarily

served by imposing on such persons results not mandated as to competent persons similarly situated."[54]

The Centrality of Quality-of-Life Judgments

Controversies about quality-of-life judgments. Our arguments thus far give considerable weight to quality-of-life judgments in determining whether treatments are optional or obligatory. When quality of life is sufficiently low that an intervention produces more harm than benefit for the patient, it is justifiable to withhold or to withdraw treatment. Such judgments require justified criteria of benefits and burdens, so that quality of life is not reduced to arbitrary judgments of personal preference and the social worth of the patient.

In a landmark case involving quality-of-life judgments, sixty-eight-year-old Joseph Saikewicz, who had an IQ of 10 and a mental age of approximately two years and eight months, suffered from acute myeloblastic monocytic leukemia. Chemotherapy would have produced extensive suffering and possibly serious side effects. Remission under chemotherapy occurs in only thirty to fifty percent of such cases and typically only for between two and thirteen months. Without chemotherapy, Saikewicz could be expected to live for several weeks or perhaps several months, during which he would not experience severe pain or suffering. In not ordering treatment, the lower court considered "the quality of life available to him [Saikewicz] even if the treatment does bring about remission." The supreme judicial court of Massachusetts, however, rejected this formulation if construed to equate the value of life with a measure of the quality of life—in particular, with Saikewicz's lower quality of life because of mental retardation. The court interpreted "the vague, and perhaps ill-chosen, term 'quality of life' . . . as a reference to the continuing state of pain and disorientation precipitated by the chemotherapy treatment."[55] It thus balanced prospective benefit against pain and suffering, finally determining that the patient's interests supported a decision not to provide chemotherapy. From a moral standpoint, we agree with the reasoning and conclusion reached in this legal opinion.

Slogans such as "quality of life" sometimes are more misleading than illuminating and thus need careful analysis. Some writers propose that we reject *moral* judgments about quality of life and rely exclusively on *medical* indications for treatment decisions. Paul Ramsey, for example, argues that for incompetent patients we need only determine which treatment is medically indicated to know which treatment is obligatory and which is optional. For dying patients, responsibilities are not fixed by obligations to provide treatments that serve only to extend the dying process, but rather by obligations to provide appropriate care in dying. The choices are thus between further palliative treat-

ments and no treatments. Ramsey worries that unless we use these guidelines we will gradually move toward a policy of active, involuntary euthanasia for unconscious or incompetent, nondying patients, based on quality-of-life judgments.[56]

However, putatively objective medical factors—such as general criteria used to determine medical indications for treatment—cannot provide what Ramsey envisions. It is impossible to determine what will benefit a patient without presupposing some quality-of-life standard and some conception of the life the patient will live after a medical intervention. Good examples are patients in a condition of permanent unconsciousness, such as Karen Ann Quinlan and Nancy Cruzan. Accurate medical diagnosis and prognosis are indispensable, but a judgment about whether to use life-prolonging measures rests unavoidably on the anticipated quality of life. The benefits of life-prolonging treatment to a permanently unconscious patient are often so limited as to render the treatment optional.[57] Unless the maintenance of mere biological life is a benefit, any benefit to the patient would appear to reside in the possibility of a diagnostic or prognostic error or of a medical breakthrough, rather than in the quality of the life that is prolonged.

Ramsey has objected that a quality-of-life approach wrongly shifts the focus from whether treatments are beneficial to patients to whether patients' lives are beneficial to them. Questioning the latter, he insists, opens the door to active, involuntary euthanasia.[58] But the principal issue is whether criteria of quality of life can be stated with sufficient precision and cogency to avoid the dangers envisioned by wedge arguments. We think they can, but the vagueness surrounding terms such as *dignity* and *meaningful life* is a cause for concern, and cases in which seriously ill or disabled newborn infants have been "allowed to die" under questionable justifications provide a reason for caution.

Several conditions of patients should be excluded from consideration. For example, mental retardation is irrelevant in determining whether treatment is in the patient's best interest. Quality of life for the patient should also not be confused with the quality or the value of life for others, and proxies should not refuse treatment against the incompetent patient's interests to avoid burdens to the family or costs to society. The incompetent patient's best medical interests generally should be the decisive criterion for a proxy, even if these interests conflict with familial interests.

The President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research recognized a broader conception of best interests that includes the welfare of the family: "The impact of a decision on an incapacitated patient's loved ones may be taken into account in determining someone's best interests, for most people do have an important interest in the well-being of their families or close associates."[59] True, a patient sometimes has an interest in the family's welfare, but it is a long step from this

premise to a conclusion about whose interests should be overriding. When the incompetent patient has never been competent or never expressed his or her wishes while competent, it is not proper to impute altruism—a desire to relieve the family of its burdens—to that patient against his or her medical best interests.[60]

Children with serious illnesses or disabilities. Some of the most difficult questions about quality of life and treatment omission involve endangered near-term fetuses, seriously ill newborns, and young children. Prenatal obstetric management and neonatal intensive care can now salvage the lives of many anomalous fetuses and disabled newborns with physical conditions that would have been fatal two decades ago. However, the resultant quality of life is sometimes so low that it raises questions about whether the aggressive obstetric management or intensive care has produced more harm than benefit for the patient. Some commentators argue that avoidance of harm (including iatrogenic harm) is the best guide to decisions on behalf of fetuses and infants in neonatal nurseries,[61] whereas others argue that aggressive intervention violates the obligation of nonmaleficence if any one of three conditions is present: "inability to survive infancy, inability to live without severe pain, and inability to participate, at least minimally, in human experience."[62]

We accept the conclusion that nonaggressive management of high-risk pregnancies and allowing seriously disabled newborns to die are under some conditions morally permissible actions, because they do not violate obligations of nonmaleficence and satisfy other justifying conditions. When quality of life is so low that aggressive intervention or intensive care produces more harm than benefit for the patient, it is justifiable to withhold or to withdraw treatment from fetuses, newborns, or infants with a variety of problems. These problems include a number of antenatal conditions that commonly eventuate in stillbirth, severe brain damage caused by birth asphyxia, Tay-Sachs disease, which involves increasing spasticity and dementia and usually results in death by age three or four, and Lesch-Nyhan disease, which involves uncontrollable spasms, mental retardation, compulsive self-mutilation, and early death. In severe cases of neural tube defects, newborns lack all or most of the brain, and death is inevitable. More problematic is meningomyelocele (protrusion of part of the covering and substance of the spinal cord because of a defect in the vertebral column). The wide range of possible outcomes from this condition makes it difficult to know whether to treat vigorously. Some children can have a meaningful life, while the chances are slim for others.

The debate about treatment or nontreatment of seriously ill and disabled newborns was stimulated in the United States by a 1973 article in which Raymond S. Duff and A. G. M. Campbell reported that forty-three of two-hundred ninety-nine consecutive deaths in the intensive care nursery at the Yale-New

Haven Hospital had occurred following a decision for nontreatment based on the infants' extremely poor prognosis for meaningful life.[63] This and similar reports led to a public debate that went unaccompanied by government intervention for almost a decade. Vigorous government action then occurred in response to the Infant Doe case, in which Infant Doe died six days after he was born with Down syndrome and respiratory and digestive complications requiring major surgery, which his parents refused to authorize. Subsequently, Congress passed amendments to the Child Abuse and Treatment Act that defined as child abuse the "withholding of medically indicated treatment" from children.[64] This law and subsequent regulations define "medically indicated treatment" as all treatment that is likely to ameliorate life-threatening conditions, including nutrition and hydration. However, three conditions are recognized under each of which life-sustaining treatment is optional.

1. The infant is chronically and irreversibly comatose
2. Provision of such treatment would merely prolong dying or not be effective in ameliorating or correcting the infant's life-threatening conditions
3. Provision of such treatment would be futile and the treatment would be inhumane

This approach has been interpreted by some influential government officials as involving reasonable *medical* judgments, rather than *quality-of-life* judgments. This strategy attempts to keep judgments in line with sound professional practice, but it is problematic on several grounds. We have already argued that "medically indicated treatments" themselves presuppose values and, often, standards of quality of life. Conditions 1 through 3 cannot be reduced to nonevaluative, medically indicated exceptive conditions to otherwise medically indicated treatments. Rather, these conditions express Congress's view of ethically indicated exceptions, and they incorporate quality-of-life judgments about which lives should be saved.[65] The judgment in condition 1 that a human life in an irreversible coma need not be prolonged is a quality-of-life judgment, and the conclusion in condition 3 requires consideration of the inhumaneness of treatment in relation to the limited prospects of success.

Consistent with our arguments in Chapter 3, the most appropriate standard in cases of never-competent patients, including seriously ill newborns, is that of best interests, as judged by the best obtainable estimate of what reasonable persons would consider the highest net benefit among the available options. Such quality-of-life judgments need to be restricted by justifiable criteria of benefits and burdens, so that quality of life is not reduced to arbitrary and partial judgments of personal preference or of the social worth of a child. For example, as we suggested previously, Down syndrome is not by itself a sufficient reason to allow a newborn to die, and usually it is not sufficient when the newborn suffers from other life-threatening conditions that require treatment.

We conclude that controlled quality-of-life considerations, together with the principle of respect for autonomy for competent patients, can legitimately determine whether treatments are optional or obligatory. These categories should replace the traditional distinctions and rules considered earlier in this chapter. However, we must now consider the most difficult of all distinctions used to determine acceptable decisions about treatment and acceptable forms of professional conduct.

Killing and Letting Die

A persistent body of distinctions and rules about life-sustaining treatments derives from the distinction between killing and letting die (or allowing to die), which in turn draws on the act–omission and active–passive distinctions. The killing-letting die distinction also underlies distinctions (1) between suicide and forgoing treatment and (2) between homicide and natural death. These distinctions are unsatisfactory for many of the purposes to which they have been put, and we will again suggest replacing some of them with categories such as benefit–burden and obligatory–optional.

However, unlike the distinctions previously discarded, we do not recommend complete abandonment of this group of distinctions. Considerations of public policy argue against a full-scale rejection. Otherwise, we risk inadequate protections for vulnerable patients, for health professionals who care for patients, and for other social groups. More than other distinctions we have criticized, distinctions centering on killing and letting die deserve some place in our moral scheme. At the same time, these distinctions are vague and need significant reformulatation, both in biomedical ethics and in public policy. To achieve an adequate moral resolution of these issues requires appeals to both beneficence and nonmaleficence that specify precisely what constitutes a harm, what constitutes a benefit, and how they are to be balanced.

Four questions need to be addressed. (1) "What conceptually is the difference between killing and letting die?" (2) "Is forgoing life-sustaining treatment sometimes a form of killing, and if so, is it sometimes suicide and sometimes homicide?" (3) "Is killing in itself no different morally than allowing to die?" (4) "Under what conditions, if any, is it permissible for patients, health professionals, or surrogates to forgo treatment so that the patient dies, to arrange for assisted suicide, or to arrange for some other cause of death?" We consider these questions in this order.

Conceptual Differences between Killing and Letting Die

Can *killing* and *letting die* be defined so that they are conceptually distinct and without overlap? The following case illustrates the problem: A newborn with Down syndrome needed an operation to correct a tracheoesophageal fistula.

The parents and physicians maintained that survival was not in this infant's best interests and decided to let the infant die rather than perform an operation. They did not consider this *omission* of treatment an act of killing the infant. However, a public outcry occurred over the case, and critics charged that the parents and physicians had killed the child by negligently allowing the child to die.[66]

In such cases, can actions that involve intentionally not treating a patient legitimately be described as "allowing to die" or "letting die," rather than "killing"? Do at least some of these actions involve both killing and allowing to die? Is "allowing to die" a euphemism for "killing"? These conceptual questions have moral implications. Unfortunately, both ordinary discourse and legal concepts are as misleading as they are helpful in our effort to understand these concepts. In ordinary language, *killing* is any form of deprivation or destruction of life, including animal and plant life. It can even mean "bringing an end to something," as in the expression "killing a legislative bill." Neither in ordinary language nor in law does the word *killing* entail a wrongful act or a crime. Ordinary language also does not require an intentional action for killing; it permits us to say that in automobile accidents one driver killed another even when no awareness, intent, or negligence was present.

Killing represents a family of ideas whose central condition is direct causation of another's death, whereas *allowing to die* represents another family of ideas whose central condition is intentional avoidance of causal intervention so that a disease or injury causes a natural death. Nevertheless, an emotive connotation of moral wrongness commonly accompanies *killing*, even under conditions that are widely considered to warrant killing, such as killing in war, self-defense, and capital punishment. This emotive connotation does not similarly affect "letting die." A need exists, then, to sharpen these notions by stipulating more precise meanings for medical ethics.

As we will use these terms, *killing* and *letting die* are properly used only to a circumstance in which one person intentionally causes the death of another human being. Killing and letting die do not occur by accident, chance, mishap, and the like. Killing and letting die also are not mutually exclusive concepts. One person can kill another by intentionally allowing the other to die, and killing can occur by omission just as well as by commission. These more precise meanings do not derive from raw stipulation on our part. Law, medicine, ethics, and ordinary language all recognize that some forms of allowing to die constitute acts of killing. As the Supreme Court of Washington put it, "the killing of a human being [can occur] by the act, procurement [that is, instigation, contrivance], or omission of another."[67] Both killing by omission and killing by commission can be intentional. Accordingly, if either a jailer or a physician withholds nutrition and hydration with the intention of ending a person's life, and an inmate or a patient dies as result, this omission is an act of killing.

Omitting Treatment as Sometimes Killing and Sometimes Letting Die

These conceptual observations have implications for our second question. Many writers in medicine, law, and ethics have construed forgoing of treatment under good medical and moral advice as *letting die* rather than *killing*, on grounds that an underlying disease or injury is the cause of death, not the forgoing of treatment. From this perspective, one acts nonmaleficently in allowing to die, but maleficently in killing. However, in many cases this thesis is difficult to uphold.

To see why, consider the following case, sometimes referred to as the Linares case: Rudolfo Linares detached his fifteen-month-old, near brain-dead son, Samuel, from a ventilator with the intent that he die. While doing so, the father prevented health professionals from reattaching the respirator by holding them away at gunpoint. Linares allowed his son to die, but the district attorney said he also killed Samuel and therefore brought a homicide complaint, although a grand jury refused to issue a homicide indictment. What would we think in this case if Linares had protested the indictment by saying, "I did not kill him; the balloon he aspirated killed him. I merely allowed him to die." Although the balloon played a causal role in the son's death, it certainly appears that the father also caused a loss of the life that was left in his son. If so, he seems to have killed his son.

However, "killing" is such a loathsome stigma in medicine that a coroner later ruled that the death was neither a case of killing nor a case of allowing to die. The death, opined the coroner, was "accidental," because Samuel had been dead when admitted to the hospital.[68] But Samuel had been determined to be alive (in a persistent vegetative state, but not brain dead) by competent physicians. This bizarre case appears to come down to the one category almost no one seems to have considered at the time: a *justified* killing. In a parallel case that we can construct as a hypothetical, if Mr. X came into a hospital off the street and detached a PVS patient from a ventilator, we would consider the act an *unjustified* killing—even though the act might be exactly the same act that a physician was about to perform at a family's request, making it a case of "allowing to die." Mr. X, the physician, and Mr. Linares all perform the same act of detaching the respirator; yet the terms we use to describe these acts are very different, as are our judgments of maleficence. Moreover, some acts of this type are maleficent, some nonmaleficent, and some are beneficent.

These conclusions have profound implications. Suppose that physicians and hospital officials had in the beginning done what both they and Rudy Linares had in fact wanted to do: detach the respirator and allow his son to die. This would have been a typical case in medicine of "allowing to die." Such acts are generally regarded as nonmaleficent, and often as beneficent. What the physician does in these cases is *causally* no different than what Rudy Linares did: Technology is removed and the patient dies as a result. If Linares killed

his son, and he did, then physicians who do the same thing with their patients likewise kill their patients. They cannot rightly say "We do not kill our patients, only the underlying diseases and injuries do so" any more than a jealous rival of a patient who detaches a respirator to ensure the patient's immediate demise could say, "It was the disease, not me, who killed him." Nor can they say that they did not intend the outcome of death. Whether the agents should be charged with and convicted of murder is a further matter for judicial judgment. Generally, the motives are proper in medicine, and both moral and legal justification exist for the action. But whether the motive is reprehensible or laudable, the act remains an intentional killing.

Why do we, like the coroner in the Linares case, resist this conclusion, which seems so straightforward? The history of the Conroy case (discussed in Chapter 3) helps to answer this question. An early appeals court found that removing the nasogastric tube from Claire Conroy was not merely a matter of forgoing treatment, because her death would be caused by dehydration and starvation. This court found that the patient "would have been actively killed" by a means independent of her medical condition, and so the act would amount to euthanasia.[69] Principal parts of this opinion were subsequently overruled by the New Jersey Supreme Court, which found that any medical treatment, including artificial nutrition and hydration, may in principle be withheld or withdrawn from an incompetent patient (under the legitimating circumstances we detailed in Chapter 3). The court held that nasogastric tubes were analytically indistinguishable from other forms of life-sustaining treatments such as respirators. No civil or criminal liability for killing is involved if appropriate conditions of refusal of treatment are present.[70]

One reason for this divergence of opinion is terminological. Some persons use the term *killing* as a *normative* term of maleficence, parallel to "unjustified homicide or murder." Justified acts involving the deaths of patients, therefore, logically cannot be instances of killing. They can only be cases of allowing to die. But, as we argued earlier, this approach risks conceptual confusion. Killing can be both morally and legally justified, despite its prima facie wrongness.

Courts have offered two primary explanations for why forgoing life-sustaining treatment should not be categorized as killing. The most prominent rationale rests on an account of causation. In acts of forgoing treatment, an underlying disease or injury is already present, and medical technology functions to prevent the natural course of the disease or injury. When the natural cause is released, a "natural death" occurs. To remove the technology, then, is to release natural conditions to do what they would have done before. Because disease or injury is the cause, rather than the physician's or surrogate's or patient's action, neither homicide nor assisted suicide occurs.

On this account, preexisting conditions alone cause the death, although this cannot be said if the technology is mischievously removed, because another

cause then enters the picture. In both *Quinlan* and *Conroy*, for example, the New Jersey Supreme Court held that the respirator is only delaying the patient's inevitable death, which would be a "natural death" if the life-support apparatus were removed. The reigning medical and legal view is that withdrawing life-sustaining treatments in such cases is an act of allowing death to occur from preexisting conditions that have been temporarily held in check by death-delaying procedures.[71]

This account functions well to alleviate physician fears about moral blame and legal liability, but it does not cohere well with many of our ordinary beliefs about killing. If Mr. X with malice detaches a conscious and perfectly competent quadriplegic from a ventilator, he does more than release natural conditions. We could not correctly say, "he didn't kill the patient; he only allowed the patient to die." By letting the patient die, he killed—in this case murdered—the patient. Physicians, by contrast, often have a solid moral and legal basis for claiming that they are warranted in omitting treatment in identical ways. They act at the patient's request or with the patient's consent. They act under a social arrangement that encourages them to do all they can to alleviate their patient's suffering. Their motive is to meet their obligation to the patient, not to serve their own interests. Therefore, we have weighty reasons for commending rather than condemning physicians in these cases.

In the attempt to protect health care professionals from charges of killing, *value* judgments about what is (morally and legally) permissible often control our *factual* judgments about the cause of death. That is, moral judgments about justified and unjustified actions determine what constitutes "killing" and what constitutes "allowing to die from preexisting conditions," rather than the other way around.[72] This leads us to say that surrogate decisionmakers and physicians do not kill patients when they justifiably remove a life-sustaining treatment, and that patients do not kill themselves when they forgo treatment; whereas if they unjustifiably omit treatment, they do kill. "Killing" is here functioning more as a *moral* category than as a *causal* category.[73] Imagine that Karen Quinlan's father had detached her respirator against medical, moral, and legal advice and that she had died as a consequence (effectively this happened in the Linares case). Few would say, "preexisting conditions killed her, her father did not kill her." But when the father has good motives and medical, legal, and moral support for the same action, many come to identify with and justify or excuse the act. As a result, they consider the act one of letting die.

Part of the reason for moralizing the cause derives from the legal doctrine of *proximate cause*. To be a proximate, or primary, cause is to become legally liable for an outcome. These legal judgments about causation and liability are often decided by a person's obligations. If a physician has an obligation to treat, then treatment omission breaches that obligation and causes death (an act of unjustified homicide); but, if no obligation to treat exists, disease or injury

serves as the proximate cause, and the physician is off the hook of liability. Conceptually, this dance with proximate causation obscures issues in the distinction between killing and allowing to die and further confuses issues about the cause of death. The doctrine assumes a single cause or single type of cause of death, yet physicians, guardians, and courts beyond a reasonable doubt play a significant causal role in bringing about death at the time it occurs in treatment-termination cases.

Even from a legal perspective, a better account can be provided than "The preexisting disease caused the death." The better account is that legal liability should not be imposed on physicians and surrogates unless they have an obligation to provide or continue the treatment. If no obligation to treat exists, then questions of causation and liability do not arise. If the categories of obligatory and optional are primary, we have a reason for avoiding discussions about killing and letting die altogether and for focusing instead on health care professionals' obligations and problems of moral and legal responsibility.

These observations are pertinent to suicide as well as to killing. Dan Brock has pointed to some relevant connections between suicide and refusal of treatment cases:

> The judgment of a person who competently decides to commit suicide is essentially that "my expected future life, under the best conditions possible for me, is so bad that I judge it to be worse than no further continued life at all." This seems to be in essence exactly the same judgment that some persons who decide to forego life-sustaining treatment make. The refusal of life-sustaining treatment is their means of ending life; they intend to end their life because of its grim prospects. Their death now when they otherwise would not have died *is* self-inflicted, whether they take a lethal poison or disconnect a respirator.[74]

Refusals of treatment, so described, are instances of suicide whenever the agent specifically arranges the conditions to bring about death. However, Brock's proposal fails to resolve the issue of whether suicidal intent and causation exist when a patient refuses treatment because of a bleak future, and this should give us pause about using the category of "suicide," as well as categories such as "killing" and "homicide." Consider a failing dialysis patient for whom the treatment has become futile: With or without treatment he is going to die in the next month. He cancels scheduled dialysis appointments in order to be at home with loved ones and die free of the machine and the hospital. This case seems to be a case lacking suicidal intent. The patient is only choosing to die under one set of circumstances rather than another. No decision is made about whether life is worth living; either way, there will soon be no life. Suicide is the wrong category, because death will, either way, be caused by untreatable conditions that are not specifically arranged by the agent for the purpose of bringing about his or her death. This is what we might call a "pure" refusal case that lacks all suicidal intent.

We conclude that the distinction between killing (suicide, homicide, etc.) and letting die suffers from vagueness and confusion. It is conceptually impossible to classify many acts as instances of letting die without also classifying them as instances of killing. We have also seen that the language of killing is so confusing—causally, legally, and morally—that we should avoid it in discussions of euthanasia and assistance in dying. It is often morally and conceptually more satisfactory to discuss these issues exclusively in the language of optional and obligatory treatments, dispensing altogether with *killing* and *letting die*.

When Killing Is No Different Morally Than Allowing to Die

Our third question, "Is killing in itself no different morally than allowing to die?" can now be addressed. To assert (as we do) that killing is no different morally than allowing to die is simply to say that correct labeling of an act as "killing" or as "letting die" does nothing to determine if one form of action is better or worse, or more or less justified, than the other. Some particular instance of killing (a brutal murder, say) may be worse than some particular instance of allowing to die (forgoing treatment for a patient who is in a persistent vegetative state, say); but some particular instance of letting die (not resuscitating a patient who could be saved, say) also may be worse than some particular instance of killing (mercy killing at the patient's request, say). Nothing about either killing or allowing to die entails judgments about actual wrongness or rightness, or about the beneficence or nonmaleficence of the action. Rightness and wrongness depend on the merit of the justification underlying the action, not on the type of action it is. Neither killing nor letting die, therefore, is per se wrongful, and in this regard they are to be distinguished from murder, which is per se wrongful. Both killing and letting die are prima facie wrong, but can be justified under some circumstances.

It would be absurd to accept all cases of letting die as morally justified, and it is no less absurd to view all forms of killing (for example, killing in self-defense) as unjustified. A judgment that an act of either killing or letting die is justified or unjustified entails that something else be known about the act besides these characteristics. We need to know something about the actor's motive (whether it is benevolent or malicious, for example), the patients' request, and the consequences of the act. Only these additional factors will allow us to place the act on a moral map and allow us to make a normative judgment about it. All instances of killing and letting die, then, must satisfy independent criteria, such as the balance of benefits over burdens to the patient, to determine their acceptability.

The Scope of the Patient's Rights

We can now address the fourth basic question, which is formulated without the language of "killing:" Under what conditions, if any, is it permissible for

patients, health professionals, or surrogates either to forgo treatment so that the patient dies or to arrange for assistance in dying?

If competent patients have a legal and moral right to refuse treatment that involves health professionals in implementing their decision and bringing about their deaths, we have a reason to suppose they have a similar right to request the assistance of willing physicians to help them control the conditions under which they die. Assuming that omission of treatment is justified by the principles of respect for autonomy and nonmaleficence, cannot the same form of justification be extended to physicians prescribing barbiturates needed by seriously ill patients, and possibly to physician-administered lethal injections? This strategy rests on the premise that reform in ethics and law is needed because of the apparent inconsistency between (1) the strong rights of autonomy that allow persons in grim circumstances to refuse treatment so as to bring about their deaths and (2) the apparent denial of a similar autonomy right to arrange for death by mutual agreement between patient and physician under equally grim circumstances.

The argument for this view seems particularly compelling when a condition has become overwhelmingly burdensome for a patient, pain management is inadequate, and only a physician can and is willing to bring relief. At present, medicine and law are in the awkward position of having to say to such patients, "If you were on life-sustaining treatment, you would have a right to withdraw the treatment and then we could let you die. But since you are not, we can only give you palliative care until you can die a natural death, however painful, undignified, and costly." This seems tantamount to condemning the patient to live a life he or she does not want.

Only a small percentage of patients face overwhelming pain and burdens because pain management and improvements in patient's environments have made circumstances at least bearable for most patients, and hospice environments have improved the care of the dying. The right to refuse nutrition and hydration also gives many patients the opportunity to control the time of their death. However, these facts do not provide a decisive reason for prohibiting increased physician assistance in dying. Some patients cannot be satisfactorily relieved, and in any event there are significant questions about autonomy rights for patients. If a right exists to stop a machine that sustains life, through an arrangement involving mutual agreement with a physician, why is there not the same right to stop the machine that *is* one's life by an arrangement with a physician?

This right has almost never been recognized in law or in codes of medical ethics. The traditional belief is that we should altogether prohibit such forms of assistance in health care while authorizing letting die in a certain range of cases. Codes of health care ethics from the time of the Hippocratic oath to the present strictly prohibit direct assistance in death, even if a patient has good

reasons for wanting to die. For example, in 1991, the American Geriatrics Society opposed all physician involvement in killing or assistance in suicide.[75] In an influential statement passed in 1973 and revised in 1988 and 1991, the American Medical Association Council on Ethical and Judicial Affairs allowed forgoing life-sustaining treatments but prohibited any "intentional termination of the life of one human being by another—mercy killing." Whether letting particular patients die is morally acceptable depends on several factors in this policy, but if the deaths involve killing—even in circumstances identical to those in which a patient is allowed to die—they are never justifiable.[76]

We have already seen that the *conceptual* distinction between killing and letting die cannot bear the weight of *normative* conclusions about policies or particular cases. Many people inside and outside medicine now believe that active physician assistance for a narrow group of seriously ill and dying patients at their request can be morally justified. Many also believe that, under closely monitored supervision, such acts of assistance in dying should be made legally permissible. As a result, an increasing number of health professionals and figures in medical ethics argue that we should relax or modify our stringent rules and laws against physician involvement.

Against this pressure for reform, many health professionals insist that practices of killing patients are inconsistent with the roles of nursing, caregiving, and healing, would introduce conflicts of interest into those roles, and would taint the roles in the same way that injecting prisoners on death row taints physicians. We need, then, to assess the arguments for and against what is often called mercy killing, assisted suicide, and omitting life-sustaining treatment with the intention of causing death. Because we earlier reached a settled view on the right to forgo treatment that generally accords with contemporary medical practice, codes of ethics, and judicial decisions, we will now consider only issues of mercy killing and assisted suicide.

Our argument will proceed as follows: First, we will present a revised account of the killing–letting die distinction in which protection against certain forms of wrongfully caused death are at the center of the discussion. We then argue that merciful physician interventions in the form of voluntary active euthanasia are not inherently wrong or incompatible with the role of a health professional. Nonetheless, public policies that sanction such physician activities are unacceptable unless they are accompanied by extraordinarily careful regulation and monitoring. Second, we argue that prohibitions in biomedical ethics against certain forms of assisted suicide should be eased, making physicians more comfortable in helping certain patients achieve what for them is a comfortable and timely death. It seems likely that assisted suicide will be the driving force behind efforts to alter rules against killing in medicine, so that support of assisted suicide, which is compatible with a rejection of voluntary active euthanasia, takes on a special significance at the present time.

Recasting the Rules Governing Physician-Assisted Death

The phrase "assisted death," particularly "physician-assisted death," is now widely used, but it is ambiguous because many modes of assistance exist. Both assisted suicide and voluntary active euthanasia are instances of assistance in bringing about death. In assisted suicide, the final agent is the one whose death is brought about, and in voluntary active euthanasia the final agent is another party. At present, voluntary active euthanasia is illegal in the United States, despite some efforts for reform. Although acts of suicide or attempted suicide have been decriminalized throughout the United States, most jurisdictions continue to prohibit aiding and abetting suicide. In one early U.S. case, a woman was bedridden with advanced multiple sclerosis and asked her husband to put a cup of poison by her bed so that she could kill herself. When she consumed the poison and died, he was prosecuted and convicted of murder on grounds that he had assisted a suicide.[77] However, many acts of assisted suicide are not prosecuted. For example, to cite a case discussed in Chapter 5 (p. 287), cancer-victim Ida Rollin indicated to her daughter, Betty Rollin, that she wanted to commit suicide, and the daughter then secured the necessary pills. Although this case was reported in newspapers, a book, and a television movie, the daughter was not prosecuted.

Some writers have argued that physician-assisted deaths are unacceptable violations of nonmaleficence, whereas others have argued that many acts of this description are acceptable and even courageous actions of beneficence. In assessing these arguments, it pays dividends to focus on the scope of the claim being advanced. We believe that sufficient moral reasons exist in some cases to justify mercy killing and assisted suicide, but these reasons are not necessarily sufficient to support revisions in either codes of ethics or public policies. In addressing whether we should retain or modify some current prohibitions, we therefore need to be clear about whether the topic of discussion is the moral justification of individual acts or the justification of institutional rules and public laws governing practices.

Acts and practices.[78] To justify an act is distinct from justifying a practice or a policy. A rule of practice or a public policy that prohibits active killing in medicine may be justifiable, even if it excludes some particular acts of causing a person's death that in themselves are morally justifiable. For example, such a rule would not permit us to use a drug overdose to cause death for a patient who suffers from terrible pain, who will probably die within three weeks, and who rationally asks for a merciful assisted death, although in an individual case the act would be justified. For policy reasons, it is sometimes necessary to prohibit such acts altogether, although they are not morally wrong.

The problem is that a practice or policy that allows killing runs risks of abuse

and, on balance, might cause more harm than benefit. The argument is not that serious abuses will occur immediately, but that they will grow incrementally over time. Society might start by severely restricting the number of patients who qualify for assistance in dying, but these restrictions would later be revised and expanded to include cases of unjustified killing. Unscrupulous persons would learn how to abuse the system, just as they do now with methods of tax evasion that operate on the margins of the system of legitimate tax avoidance. In short, the slope of the trail toward unjustified killing will be so slippery and precipitous that we ought never to hike on it.

Questions about the slippery slope. Many dismiss these slippery slope or wedge arguments because of their widespread abuse in biomedical ethics, a lack of empirical evidence to support their claims, and their heavily metaphorical character ("the thin edge of the wedge," "the first step on the slippery slope," "the foot in the door," and "the camel's nose under the tent"). However, some arguments of this form should be taken with the utmost seriousness.[79] They force us to think carefully about whether unacceptable harm is likely to result from attractive and apparently innocent first steps.

Wedge or slippery slope arguments appear in two versions: (1) conceptual and (2) psychological–sociological. According to the first version, the slope is slippery because the concepts and distinctions used in moral and legal rules are vague and may lead to unanticipated outcomes. A norm or justification for a type of action that, considered in isolation, is morally acceptable winds up supporting similar acts that are unacceptable. For example, some justifications that have been offered for the moral acceptability of suicide imply a justification of some forms of voluntary active euthanasia that seem unjustifiable to proponents of the initial justification. Critics then argue that the justification that was offered for suicide is the first step on the slippery slope or the thin edge of the wedge toward voluntary, active euthanasia. This first version of the wedge or slippery slope argument, however, can also be used against its proponents. If it is morally defensible to allow patients to die under conditions *x*, *y*, and *z*, then (in light of our previous argument) it is morally defensible to assist them more aggressively in bringing about their deaths under those identical conditions. If it is in their interests to die, it is (prima facie) irrelevant how death is brought about.

The second or psychological–sociological version of slippery slope arguments offers a better reason for maintaining the distinction between killing and letting die. This version examines the probable impact of making exceptions to professional, social, and legal rules or changing them in a more permissive direction. If certain restraints against killing are removed, various psychological and social forces would likely make it more difficult to maintain the relevant distinctions in practice. For example, in some settings it is plausible to

argue as follows: (1) To authorize killing patients for their benefit when they are suffering excruciating pain or have a bleak future risks opening the door to the encouragement of euthanasia in order to relieve personal burdens on families and in financial burdens on society. (2) *Voluntary* active euthanasia (an act of killing a person at his or her informed request) invites social changes leading to *nonvoluntary* euthanasia (an act of killing a person who is incapable of making an informed request) and perhaps to *involuntary* euthanasia (an act of killing a person who while competent opposes being killed). In assessing these possibilities, we should recall that killing and various forms of assistance in dying can occur by both omission and commission. Withheld or withdrawn treatment (such as hydration and nutrition) can cause death just as an underlying disease or injury can cause death.

This second version of the slippery slope argument becomes more compelling when we consider the effects of discrimination based on disability, the increasing number of newborns with disabilities who survive at heavy cost to the public, and the growing number of aging persons with medical problems who require larger and larger proportions of the public's financial resources. If rules permitting voluntary active euthanasia become public policy, society is at increased risk that persons in these populations will be harmed; for example, the risk is increased that families and health professionals may kill disabled newborns and severely brain-damaged adults to avoid social and familial burdens. If newborns and adults can be judged by decisionmakers to have overly burdensome conditions or lives with no value, the same logic can be extended to many other populations of feeble, debilitated, and seriously ill patients who are financial and emotional burdens on families and society.

Many of these circumstances are relevantly similar to circumstances that already provide the leading justifications for widely accepted forms of withdrawing or withholding life support. Often the patients did not request these omissions and left no advance directive. These cases differ by degree as much as by kind, which makes it easier to extend the same reasoning to other cases. It takes little imagination to suppose that many parents would, if given the opportunity, withhold life-sustaining technologies from their newborns because of a wide variety of disabilities, such as blindness, retardation, and malformed limbs.

Rules in our moral code against actively causing the death of another person are not isolated fragments. They are threads in a fabric of rules that support respect for human life. The more threads we remove, the weaker the fabric becomes. If we also focus on the modification of *attitudes*, not *rules* only, the general attitude of respect for life can also be eroded by shifts in public policy. Prohibitions are often both instrumentally and symbolically important, and their removal could weaken a set of practices, restraints, and attitudes that we cannot replace.[80]

Rules against bringing about another's death also provide a basis of trust between patients and health care professionals. We expect health care professionals to promote our welfare under all circumstances. We risk a loss of public trust if physicians become agents of active euthanasia in addition to healers and caregivers. At the same time, we risk a loss of trust if patients and families believe they are being abandoned in their suffering by physicians who lack the courage and will to offer the assistance they believe they need in the darkest hours of their lives.

The ultimate success or failure of these slippery slope arguments depends on speculative predictions of a progressive erosion of moral restraints. If dire consequences will in fact flow from the legal legitimation of assisted suicide or voluntary active euthanasia, then the argument is cogent and such practices are justifiably prohibited. But how good is the evidence that dire consequences will occur? Does the evidence indicate that we cannot maintain firm distinctions in public policies between patient-requested death and involuntary euthanasia? Scant evidence supports any of the answers that have been given to these questions, so far as we can see. Those, including the present authors, who take seriously the second version of the slippery slope argument should simply admit that the argument needs a premise on the order of "better safe than sorry." The likelihood of the projected moral erosions, then, is not something we can easily assess. Arguments on every side are speculative and analogical, and different assessors of the same evidence reach different conclusions. An intractable controversy also exists over what counts as good evidence. Although we cannot here resolve this largely empirical issue, we can assess an analogy that frequently arises in these discussions: the Nazi path to a final solution.

The Nazi analogy. The holocaust continues to serve as a powerful vision of the bottom of the slippery slope for a society that carelessly initiates killing. The holocaust left a string of inadequately answered questions about so-called euthanasia. After the Nuremberg trials of German physicians, American physician Leo Alexander argued that the Nazis moved from the "small beginnings" of euthanasia for the incurably ill to policies of genocide:

> The beginnings at first were merely a subtle shift in emphasis in the basic attitude of the physicians. It started with the acceptance of the attitude, basic in the euthanasia movement, that there is such a thing as life not worthy to be lived. This attitude in its early stages concerned itself merely with the severely and chronically sick. Gradually the sphere of those to be included in this category was enlarged to encompass the socially unproductive, the ideologically unwanted, the racially unwanted and finally all non-Germans. The infinitely small wedged-in lever from which this entire trend of mind received its impetus was the attitude toward the nonrehabilitable sick.[81]

This account reappears in Robert Lifton's study of Nazi physicians, which describes the first steps as well as the final horror of the rule that "life unworthy of life" is to be eliminated.[82] Lifton notes that prior to the death camps, the

Nazis adopted a policy of direct medical killing, using injections, lethal doses of medicine, and gases. The killing was arranged within the medical system and involved medical decisionmakers. Doctors and their assistants implemented the decisions. Crucial to the program was the removal of a social and psychological barrier against killing through a "medicalization of killing" that obscured the boundaries between killing and helping. Lifton argues that although this program was called *euthanasia*, eventually the term simply "camouflaged mass murder."

Contemporary proponents of euthanasia often properly insist that the rationale of the Nazi program was racist ideology, not respect for autonomy and traditional values in health care. They dispute the Nazi analogy, because the Nazis concentrated on nonvoluntary and especially involuntary killing (inappropriately labeled *euthanasia*), and did not innocently step onto the slippery slope only to find they could not stop.[83] We accept this argument that the Nazi analogy is weak and that mercy killing is not always wrong. At the same time, society must protect its members against disastrous outcomes by designing appropriate social policies and statements of professional ethics that prevent abuse.

An example of reckless mercy killing that prohibitory rules should help deter was reported in the *Journal of the American Medical Association* in January 1988 under the provocative title "It's Over, Debbie."[84] A gynecology resident rotating through a large private hospital was awakened by a telephone call from a nurse who told him that a patient on the gynecologic-oncology unit, not the resident's usual duty station, was having difficulty getting rest. The chart at the nurses' station provided some details. A twenty-year-old woman named Debbie, who was dying of ovarian cancer, was experiencing unrelenting vomiting from the alcohol drip administered for sedation (a procedure that some have criticized). The woman was emaciated, weighed eighty pounds, had an intravenous line, was receiving nasal oxygen, and was sitting in bed suffering from severe air hunger. She had not eaten or slept in two days, and she was receiving only supportive care because she had not responded to chemotherapy. The patient's only words to the resident were, "Let's get this over with." After having the nurse draw twenty milligrams of morphine sulfate into a syringe, the resident took it into the room and injected it intravenously into the patient after telling her that it "would let her rest" and "say good-bye." The patient died within a few minutes.

If this is an actual case—and doubts have been voiced about its authenticity—the resident acted rashly. Other medications could perhaps have relieved the patient's pain and suffering and enabled her to rest comfortably. The resident's intention seems to have been to kill the patient out of "mercy," but in the absence of previous contact with the patient, the resident had no basis for interpreting her words as a request to be killed and did not consult with anyone

before making a quick, momentous, and irreversible decision. Both law and ethics should deter such actions.

Merciful death in the practice of medicine. In addition to fears of abuse of individuals such as the physically and mentally disabled who cannot consent, other legitimate fears haunt active medical interventions to bring about death. Consider the following two types of wrongly diagnosed patients:[85]

1. Patients who are wrongly diagnosed as hopeless and who will survive if a treatment is ceased (in order to allow a natural death)
2. Patients who are wrongly diagnosed as hopeless and who will survive only if the treatment is *not* ceased (in order to allow a natural death).

If a social rule that allows some patients to die were in effect, doctors and families who followed it would lose patients only in the second category. But if killing were permitted, at least some of the patients in the first category would be needlessly lost. Thus, a rule prohibiting killing would save some lives that would be lost if both killing and allowing to die were permitted. This consequence is not a decisive reason for a policy of (only) allowing to die, because the numbers in categories 1 and 2 are likely to be small, and other reasons for assisting in dying, such as extreme pain and autonomous choice, might be weighty. But it is a morally relevant reason for a cautious policy that calls for careful review and monitoring of decisions.

Among the strongest reasons for mercifully helping some patients to die, of course, is the relief of unbearable and uncontrollable pain and suffering, which can so ravage and dehumanize patients that death appears to be in their best interests. Prolonging life and refusing to kill in some of these circumstances seems a cruel violation of the principle of nonmaleficence by causing people to suffer. Nevertheless, an array of alternatives can be presented to many of these patients. The physician can usually relieve pain and make a patient comfortable through medications, even if the medications hasten death. Physicians can also appropriately, and often painlessly, withdraw nutrition and hydration. In many cases the patient will view this action as the best alternative. One reason for a policy of exhausting all alternatives before allowing physicians to engage in active euthanasia is the precariousness of constructing a social or professional ethic on borderline situations and emergency cases. It is dangerous to generalize from emergencies, because hard cases may make bad social and professional ethics as well as bad law.

Clinicians also have a moral obligation to inform competent patients of alternative approaches, such as hospice care and increased medication. The risk of addiction has often been overestimated and unduly feared in the care of terminally ill patients.[86] Public policy in the United States has resisted legalizing

heroin, a powerful painkiller. Government officials fear the harmful consequences that might flow from such an act, including addiction, the legitimation of heroin, and the possibility of abuses. Yet heroin has been used for terminally ill cancer patients for several years in the United Kingdom without evidence of major problems. We thus see no merit in the societal prohibition of the use of heroin to relieve pain in terminally ill cancer patients.

An ongoing controversy surrounds one experiment in socially accepted, voluntary active euthanasia—the case of the Netherlands, where euthanasia is still technically illegal, but where guidelines developed by courts also immunize physicians against prosecution. Euthanasia is openly practiced, and it is supported by a substantial segment of the population. Defenders of euthanasia contend that the experiment in the Netherlands establishes that euthanasia can be socially accepted without the parade of horrors that many people have predicted. But critics contend that the Netherlands fails as a model for the United States for at least two reasons. First, social conditions in the United States are vastly different from those in the Netherlands, where a more homogeneous population carries universal access to health care and typically has close relations to primary care providers. Second, some physicians and families in the Netherlands have gone beyond the accepted rules by putting to death some incompetent patients, such as infants with Down syndrome, although the rules authorize only voluntary euthanasia. According to one nationwide study in Holland, about 1.8 percent of the total deaths each year are the result of euthanasia by physician-administered lethal drugs at the patient's request, and about 0.3 percent are the result of physician-assisted suicide; in 0.8 of all deaths, drugs are administered to shorten a patient's life "without explicit and persistent request" and without satisfying the nation's stringent criteria for euthanasia.[87]

Although all of the above arguments deserve our deepest respect, they do not answer every question that needs to be addressed about physician-assisted suicide and mercy killing, for at least four reasons. First, we are already on a slippery slope by virtue of the changes in professional ethics and law that have occurred since *Quinlan* and the first natural death acts. If it is morally permissible to unplug respirators and detach intravenous lines knowing that death will eventuate, the logic of our present situation is that we are struggling to preserve as many traditional restraints against killing as we can, consistent with taking a humane approach toward seriously suffering patients and respecting their rights. Second, as we have seen, arguments on every side of the debate about the slippery slope are speculative and analogical, and different assessors of the available evidence legitimately reach different conclusions. Third, law and public policy are not likely ever to be decisive guidelines for medical ethics. We are sometimes justified in performing actions that evade or infringe legal and social rules, and we should never allow medical ethics to be totally determined by sound social policy. Fourth, we often accept social policies that have some

risks, knowing that tragedies will sometimes occur. For example, we allow youths sixteen years of age to drive an automobile on public roads with minimal driver's education, although we know there will be some tragic outcomes. The rationale is that, on the whole, the results of the practice will be more beneficial for all affected than any competing policy we can fashion. Proponents of voluntary active euthanasia and patients in uncontrollable pain are now asking us to accept a similar rationale. They want a sound monitoring system under which people elect to die, even though a few tragic mistakes will result from such a system.

Is there, then, a reasonable hope that a public policy of legalized euthanasia can be properly monitored and enforced without persistent and pervasive abuse? Our apparent success in managing many cases of treatment omission that qualify as forms of passive euthanasia in health care institutions and the courts is an encouraging sign that abuses are containable, although only a weak body of data exists. There is a significant problem of inconsistency in our policies if we maintain that legalizing voluntary *active* euthanasia will lead to unacceptable practices and consequences, whereas allowing both *passive* voluntary euthanasia and termination of treatment by surrogates will not present comparable difficulties. To date, the more difficult problems have not come in cases of voluntary patient requests, where patients are available for discussion and decisionmaking, but in cases of incompetent patients, such as Earle Spring and Nancy Cruzan. However, this book is not the appropriate forum in which to work out a precise framework for legal mechanisms and public policy. We are not convinced that legalization is either the best public policy or a policy that we should reject. In any event, the basic question that confronts contemporary biomedical ethics is not the precise limits in public policy that should be placed on assisting in death (as momentous as this question is), but rather which forms of assistance in death are morally justified.

The Justification of Assistance in Dying

There are, we believe, sound reasons to accept several forms of assistance in dying that help bring about death. Some of these forms of assistance qualify as voluntary active euthanasia, others as passive euthanasia.

Why Is It Wrong to Bring about Death?

Most of us believe it is not always morally wrong, on balance, to cause someone's death. Why, then, do we believe that causing someone's death is prima facie wrong, and what could make it right?

We have encountered one answer in examining slippery slope arguments: It is wrong to kill whenever killing threatens social stability or causes harmful

social consequences. But this answer does not explain the wrongness of causing death, which has little, if anything, to do with parties other than the person whose life is taken. Causing a person's death is wrong because of a harm or loss to the person killed, not because of losses that others encounter. What makes it wrong, when it is wrong, is that a person is harmed—that is, suffers a setback to interests that the person otherwise would not have experienced. In particular, one is caused the loss of the capacity to plan and choose a future, together with a deprivation of expectable goods. This explains why inflicting death both harms and wrongs a person.

This conclusion is notable for the following reason: If a person desires death rather than life's more typical goods and projects, then causing that person's death at his or her autonomous request does not either harm or wrong the person (though it might still harm others—or society—by setting back their interests, which might be a reason against the *practice*). To the contrary, not to help such persons in their dying will frustrate their plans and cause them a loss, thereby harming them. It can also bring them indignity and despair. Furthermore, if *passive* allowing to die does not harm or wrong a patient because it does not violate the patient's rights, then assisted suicide and voluntary active euthanasia similarly do not harm or wrong the person who dies. Those who believe it is sometimes morally acceptable to let people die but not to take active steps to help them die must therefore give a different account of the wrongfulness of killing persons than the one we have suggested. The burden of justification, then, seems to rest on those who would refuse assistance to those who wish to die, rather than on those who would help them.

Justified Breaches of Legal Rules

Juries often excuse those who kill their suffering relatives by finding them not guilty by reason of temporary insanity. As we suggested in the section above on the rule of double effect, excusability is usually based on a judgment of the person's motive of mercy, not an assessment of the person's intentions. Consider a once famous case in New Jersey.[88] George Zygmaniak was in a motorcycle accident that left him paralyzed from the neck down. The paralysis was considered irreversible, and George begged his brother to kill him. Three days later, his brother brought a sawed-off shotgun to the hospital and shot George in the head after saying, "Close your eyes now, I'm going to shoot you." A judgment of temporary insanity in this case is a veiled moral judgment of acceptable killing. The judgment of "insanity" springs from the lack of a legal channel to say the act was, under the circumstances, justifiable. Verdicts such as "not guilty by reason of temporary insanity" function under law to excuse the agent by finding (sometimes implausibly) that he or she lacked the conditions of responsibility necessary for legal guilt. George Zygmaniak's brother

was not temporarily insane, but the jury found the act morally excusable nonetheless.

Physicians (and others) on occasion find that it is morally permissible to engage in justified conscientious or civil disobedience of laws against killing and assisting in dying. This is another way of acknowledging that there are justified moral exceptions to enforceable rules against killing. The conditions that justify conscientious refusals to follow rules against killing patients are too complex to be considered here, but the key point is this: If pain and suffering of a certain magnitude can in principle justify active interventions to cause death, then acts of conscientious refusal to follow laws will sometimes be justified (as long as certain other conditions are met). It may or may not be justified to introduce parallel changes in the law. The language of "conscientious refusal" is not used to evade acceptance of the justifiability of active euthanasia in the difficult cases. We accept its justifiability for reasons now to be explained. (We discuss "conscientious refusal" in Chapter 8.)

Physician-Assisted Suicide

Debates about suicide have long occurred in medical practice, particularly psychiatric practice, because physicians have usually intervened to prevent suicides and to treat patients who attempt suicide (see Chapter 5 pp. 284–287). Physicians have also assisted patients in committing suicide, despite legal and professional prohibitions. But should physicians be given a more extensive role in facilitating suicide than medical ethics and social convention have traditionally permitted?

Jack Kevorkian's first use of his now famous suicide machine offers an example of *unjustified* physician-assisted suicide that medical ethics should discourage. In his first case, Janet Adkins, an Oregon grandmother with Alzheimer's disease, had reached a decision that she wanted to take her life rather than lose her cognitive capacities, which she was convinced were slowly deteriorating. After Adkins read about Kevorkian's machine in the news media, she communicated with him by phone and then flew from Oregon to Michigan to meet with him. Following brief discussions over a weekend, she and Kevorkian drove to a park in north Oakland County. He inserted a tube in her arm and started saline flow. His machine was constructed so that Adkins could then press a button to inject other drugs, eventuating in potassium chloride, which physically caused her death.[89] She then pressed the button.

This case raises several concerns. Janet Adkins was in the fairly early stages of the crippling effects of Alzheimer's and was not yet debilitated. Her death was on the distant horizon. At fifty-four years of age, she was still capable of enjoying a full schedule of activities with her husband and playing tennis with her son, and she might have been able to live a meaningful life for several

more years. There was a slight possibility that the Alzheimer's diagnosis was incorrect, and she might have been more psychologically depressed than Kevorkian appreciated. More importantly, she had limited contact with him before they collaborated in her death, and he had not administered examinations to confirm either her diagnosis or her level of competence to commit suicide. He also lacked the professional expertise to evaluate her. The glare of media attention also raises the question whether Kevorkian acted imprudently in order to generate publicity for his suicide machine and for his forthcoming book.

Jack Kevorkian's actions have been almost universally condemned by lawyers, physicians, and writers in ethics. The case raises all the fears present in the arguments mentioned previously about killing in medicine: abuse, lack of social control, physicians acting without accountability, and unverifiable circumstances of a patient's death. Although this approach to assisted suicide is improper, Kevorkian's "patients" do raise profoundly distressing questions about the lack of a support system in medicine or elsewhere for handling their problems. Having thought for over a year about her future, Janet Adkins decided that the suffering of continued existence exceeded the benefits. Judging from her friends' reports, she knew exactly what she wanted and appreciated both the costs and the benefits. Her family supported her decision, however much they disagreed with it. She faced a bleak future from the perspective of a person who had lived an unusually vigorous life, both physically and mentally. She believed that her brain would be slowly destroyed, with progressive and devastating cognitive loss and confusion, fading memory, immense frustration, and lack of all capacity to take care of herself. She also believed that the full burden of responsibility for her care would be placed on her family. From her perspective, what Kevorkian offered was preferable to what other physicians offered.

Current social institutions, including the medical system, are inadequate to help many patients in a similar condition who have reached a similar conclusion about their fates. Many dying persons face inadequate counseling, emotional support, and pain control. To them their condition is intolerable, and no avenue of hope exists. They would rather kill themselves or be killed than face what they understand to be a bleak future without relief. To say that these persons act immorally by arranging for death at their own hand or with a physician's assistance is a harsh judgment that needs to be backed by persuasive argument. Is it, then, justifiable for physicians to assist in their suicides?

Several prominent cases of *justified* assisted suicide have emerged in medicine in recent years, despite its widespread illegality. First, consider the case of Larry McAfee, in which a court as well as physicians faced a dilemma about legitimate forms of assistance. McAfee was a competent adult paralyzed from the neck down as the result of an automobile accident. He was not terminally ill, but he found his life as a quadriplegic intolerable. A professional engineer, he devised a self-disconnecting, mouth-controlled mechanism that would separate him from his ventilator, thereby causing his death. A Georgia court found that McAfee's right to refuse treatment and disconnect himself outweighed the state's interest in the preservation of life and in preventing suicide. This finding is an endorsement of the right of a competent patient to refuse customary life-sustaining treatment.

But McAfee wanted more, from the courts and from his physicians. He had previously attempted to disconnect himself from the respirator, but had been unable to follow through with the act because he was incapacitated from loss of oxygen. He therefore asked for a physician's assistance in administering a sedative just before he attempted to disconnect himself. The court found in 1989 that no criminal or civil liability would be attached to a physician who helped him by administering the sedative, but this court hinted (agreeing with a trial court) that courts could not order a physician to administer the sedative. Nevertheless, the court found that "McAfee's right to have a sedative (a medication that in no way causes or accelerates death) administered before the ventilator is disconnected is a part of his right to control his medical treatment."[90]

These confusing and troublesome cases should never reach a court. The right acknowledged is a right that health care facilities should recognize without the patient having to meet repeated refusals of assistance from physicians. We do not propose a right that requires coercion of the conscience of physicians, a troubled area of medical ethics. We are recommending instead that medical professionals themselves confront these issues more directly and acknowledge that it is permissible to assist patients. It would then not be necessary to *require* assistance. The problem is that law and medicine (and to some extent ethics) have conspired to block this option for patients by insisting on the maintenance of traditional sanctions against physician-assisted suicide. Larry McAfee is a striking example of how the current system drives patients who need medical attention to physicians like Jack Kevorkian who are willing to take more aggressive actions.

The appellate court in the case of Elizabeth Bouvia pushed matters further. The court suggested that there is a right of privacy to commit suicide and that courts and physicians morally should make it possible for physicians to assist patients in bringing about the end of their lives in dignity and comfort. This court expressly attempted to widen the boundaries of justifiable active assistance by physicians in bringing about a patient's death. In a concurring opinion, Associate Justice Compton exhorted physicians to rethink their traditional objections to assisting such patients to die. The right to die, he said, includes the right to secure assistance from members of the medical profession.[91]

Finally, we turn to a case that many find troublesome but that we believe to be a case of justified assisted suicide. This case involves physician Timothy Quill, who prescribed the barbiturates desired by a 45–year-old patient who

had refused a risky, painful, and often unsuccessful treatment for leukemia. She had been his patient for many years, and members of her family had, as a group, come to this decision with the counsel of the physician. The patient was competent, and all reasonable alternatives for the relief of suffering had been discussed and rejected. Several of the conditions that the present authors consider sufficient for justified assisted suicide were satisfied. These conditions include:

1. A voluntary request by a competent patient
2. An ongoing patient–physician relationship
3. Mutual and informed decisionmaking by patient and physician
4. A supportive yet critical and probing environment of decisionmaking
5. A considered rejection of alternatives
6. Structured consultation with other parties in medicine
7. A durable preference for death expressed by the patient
8. Unacceptable suffering by the patient
9. Use of a means that is as painless and comfortable as possible

Even though Quill's actions satisfied most of these conditions, some people find his involvement as a physician unsettling and unjustified. The wedge argument has been mentioned, because so many patients, especially in elderly populations, are potentially affected if acts like Quill's are legalized. Others are troubled by the fact that Quill potentially violated a New York State law against assisted suicide. (After he wrote an article on the case, a grand jury in Rochester, New York, where the events occurred, declined to indict him, apparently because jurors sympathized with his motives and possibly his action.) Furthermore, to reduce the risks of criminal liability, Quill lied to the medical examiner by informing him that a hospice patient had died of acute leukemia.[92]

Despite these problems, we do not oppose Quill's act, his patient's decision, or their relationship. Suffering and loss of cognitive capacity can ravage and dehumanize patients so that death is in their best interests. In these tragic situations physicians such as Quill do not act wrongly in assisting competent patients to bring about their deaths. Public policy issues regarding how to avoid abuses and discourage unjustified acts[93] should be part of our discussion about assisted suicide, but these issues are not problems about the justifiability of the physician's act.

In general we have thus far been able to respect the line between unjustifiable and justifiable passive euthanasia in medical practice, and we should similarly be able to hold the line between justified and unjustified assistance for suicide. However, we are aware that this observation is not entirely free of conflict with our earlier comments on wedge or slippery slope arguments. We therefore need

to reconcile—that is, bring into reflective equilibrium—these two points of view.

The view we recommend is the following: Physicians have traditionally maintained that they have no obligation to assist in suicide, only an obligation to care for patients in the process of their dying and an obligation to "do no harm." This position suggests that the act of assisting, if justifiable, is never obligatory; at best, it is a merciful form of nonrequired assistance. This attitude needs to change in medicine. We need to reconceive certain forms of assisting in dying as part of the responsibility of caring for the patient, while rejecting other forms of assistance as outside that obligation. The focus of the discussion about euthanasia and assisted suicide in upcoming years should be on traditional attitudes in medicine, the policies they have generated, and ways to redraw the unstable and often indefensible lines in these policies. As these policies are reconsidered for competent patients, we will also need to reconsider policies for incompetent patients.

Decisionmaking for Incompetent Patients

We discussed standards for surrogate decisions for incompetent patients in Chapter 3. We will now consult those standards in order to discuss *who* should decide for the incompetent patient.

Typically, we think of families who care deeply about their elderly and incompetent members. However, this focus is too narrow. We need an approach that includes incompetent individuals who lack family members and the many residents of nursing homes, psychiatric hospitals, and facilities for the disabled and mentally retarded who rarely, if ever, see a family member. The appropriate roles of families and courts, guardians, conservators, hospital committees, and health professionals all merit consideration.

All we can hope for in treatment and nontreatment decisions for incompetent patients is imperfect procedural justice[94]—that is, a procedure that is just but cannot ensure or guarantee the right outcome (as judged by some independent standard). For example, in criminal trials an independent standard of a right verdict exists (conviction of the guilty and only the guilty), but it is impossible to design a procedure that will guarantee the right verdict in every case. We must, then, evaluate procedures for decisionmaking for incompetent patients according to their fairness and reliable (but imperfect) production of right outcomes.

Advance Directives

In an increasingly popular procedure rooted more in autonomy than in nonmaleficence, a person while competent either writes a directive for health care

professionals or selects a surrogate to make decisions about life-sustaining treatments during periods of incompetence.[95] Both actions are appropriate exercises of autonomy. We need, then, to distinguish two types of *advance directive* aimed at governing future decisionmaking: (1) *living wills*, which are specific substantive directives regarding medical procedures that should be provided or forgone in specific circumstances, and (2) *durable power of attorney* (DPA) for health care, or proxy directives. A DPA is a legal document in which one person assigns another person authority to perform specified actions on behalf of the signer. The power is "durable" because, unlike the usual power of attorney, it continues in effect if the signer of the document becomes incompetent.

Much of the early legislative action on this topic (principally involving natural death acts) focused on the agent's decisions through living wills, in the form of advance directives to physicians that specify the treatment a person welcomes or declines in foreseeable circumstances such as a persistent vegetative state (PVS), irreversible loss of cognitive capacities, and incompetence. However, it has proved difficult to specify decisions or guidelines that adequately anticipate the full range of medical situations that might occur, and recently the trend has been to designate surrogates. Both kinds of advance directive can be combined in some legal jurisdictions in a single document,[96] and both can be used for refusal of life-sustaining treatment.

Living wills and DPAs protect autonomy interests and may reduce stress for families and health professionals who fear making the wrong decision, but they also generate both practical and moral problems.[97] First, relatively few persons compose them or leave explicit instructions.[98] This situation is unlikely to change with increased public awareness and patient education. Second, a designated decisionmaker might be unavailable when needed, might be incompetent to make good decisions for the patient, or might have a conflict of interest (for example, because of a prospective inheritance or an improved position in a family-owned business). Third, some patients who change their preferences about treatment fail to change their directives, and a few who become legally incompetent protest a surrogate's decision. Fourth, state laws are often written to severely restrict the use of advance directives. For example, they have legal effect in some states if and only if death is imminent and the patient is terminally ill and incompetent. But decisions must be made in some cases when death is not imminent or the medical condition cannot appropriately be described as a terminal illness. Fifth, living wills provide no basis for health professionals to overturn instructions that turn out not to be in the patient's best medical interest, although the patient could not have reasonably anticipated this circumstance while competent. Surrogate decisionmakers too make decisions with which physicians sharply disagree, in some cases asking the physician to act against his or her conscience. Sixth, some patients do not have an adequate

understanding of the range of decisions a health professional or a surrogate might be called upon to make, and even with an adequate understanding it is frequently difficult to foresee clinical situations and possible future experiences.

Many living wills are phrased in vague terms, such as (to approximate AMA language cited earlier), "In the event of a terminal illness where irrefutable evidence exists that biological death is imminent, all extraordinary means of life support should be discontinued." For example, The Pennsylvania Advance Directive for Health Care "declares" as follows: "I direct my attending physician to withhold or withdraw life-sustaining treatment that serves only to prolong the process of my dying, if I should be in a terminal condition or in a state of permanent unconsciousness."[99] Additional questions often must be answered later, such as "Is this condition terminal?," "Is NG-feeding (nasogastric feeding) extraordinary means?," "Is CPR heroic?," and "Is death imminent?" Inference and discretion are involved in answering such questions. The DPA therefore is a more practical instrument than a living will.

Many of these problems can be handled through more carefully worded documents and through proper counseling and skillful explanations by physicians of medical possibilities and treatment options, but some problems of interpretation will remain despite increased physician involvement and educational tools such as videotapes. The need for resourceful interpretation is illustrated in the following case: Mrs. Z., a fifty-five-year-old teacher of foreign languages, developed aspiration pneumonia, which required admission to the intensive care unit. Her condition was probably caused by a diminished gag reflex, the result of twenty years of multiple sclerosis. To prevent future occurrence, the staff discussed oversewing the patient's epiglottis (part of the larynx), which would require a permanent tracheostomy and entail loss of laryngeal speech capability. Because of her multiple sclerosis, Mrs. Z. was confined to bed at home. Her only interaction with friends involved speech, and she tutored students at home. Without the medical procedure, a future episode of aspiration pneumonia would probably be fatal. Mrs. Z. stated that she would rather die than be unable to speak, but she was not clearly competent at the time, in part because of what was believed to be a mild organic brain syndrome. Mrs. Z.'s prior living will was then submitted by her sister. The document contained Mrs. Z.'s directive not to be kept alive artificially if she could not lead a "useful life." The sister—in effect serving as a proxy, as if there were a DPA—interpreted Mrs. Z.'s use of the phrase "useful life" to include the ability to relate to others meaningfully by verbal communication. The staff felt comfortable in accepting this judgment and in refraining from performing procedures they had been considering.[100]

Despite the questionable interpretation of "useful life" in this case, and despite the six problems cited previously, the advance directive is a promising way for competent persons to exercise their autonomy. From the perspective of

ethical theory, none of these problems is theoretically untroubled. The problems are primarily practical, and some can be overcome by adequate methods of implementation that follow the outlines of procedures for informed consent in Chapter 3.

Surrogate Decisionmaking without Advance Directives

When an incompetent patient has not left an advance directive, who should make the decision, and with whom should the decisionmaker consult?

Qualifications of surrogate decisionmakers. We propose the following list of qualifications for decisionmakers for incompetent patients (including newborns):

1. Ability to make reasoned judgments (competence)
2. Adequate knowledge and information
3. Emotional stability
4. A commitment to the incompetent patient's interests that is free of conflicts of interest and free of controlling influence by those who might not act in the patient's best interests.

The first three conditions are familiar from the discussion of informed consent in Chapter 3. The only potentially controversial condition is the fourth. Here we are endorsing a criterion of *partiality*, acting as an advocate in the incompetent patient's best interests, rather than *impartiality*, which requires neutrality in the consideration of the interests of the various affected parties.[101]

Four classes of decisionmakers have been proposed and used in cases of withholding and terminating treatment for incompetent patients: families, physicians and other health care professionals, institutional committees, and courts. If a court-appointed guardian exists, that person will be the primary responsible party. But absent the intervention of a court, we need a defeasible structure of decisionmaking authority that places the family as the presumptive authority when the patient cannot make the decision and has not previously designated a decisionmaker.

The role of the family. It is now widely agreed that the patient's closest family member is the first choice as a surrogate. The family's role should be presumptively primary because of expectable identification with the patient's interests, intimate knowledge of his or her wishes, depth of concern about the patient, and the traditional role of the family in society. However, the patient's closest family member(s) are demonstrably unsatisfactory in some cases, and the authority of the family is not final or ultimate.[102] Circumstances occur in which

physicians rightly feel compelled to reject a family's decision or to require its review by an ethics committee or the courts. Even the closest family member can have a conflict of interest, can be poorly informed, or can be too distant personally (even estranged). Challenges to family authority of course need to be supported by evidence of the potentially unreasonable or harmful character of their decisions.[103]

Unfortunately the term *family* is imprecise, especially if the extended family is included. Our reasons for assigning presumptive priority to the patient's closest family member also support assignment of priority to other family members, as most state statutes now require. The ranking varies in these statutes, but an example we find acceptable is the ordering in the Virginia Natural Death Act. If the patient is incompetent and has not specified standards through an advance directive, a decision to withhold or withdraw life-prolonging treatment must involve consultation and agreement between the attending physician and "any of the following individuals in the following order of priority if no individual in a prior class is reasonably available, willing and competent to act:" judicially appointed guardian (if necessary in the circumstances), patient-designated decisionmaker, spouse, adult child or a majority of the adult children reasonably available, parents of the patient, and nearest living relative of the patient.[104]

For a previously competent patient, this serial arrangement of family members—spouse, adult children, parents, etc.—rests on their presumed ability to use the person's expressed preferences or values to make the decision or to interpret the standard of best interests, as well as their presumed willingness to do so. For a newborn, the parents generally should be the primary decisionmakers, because they have engaged in a series of actions that resulted in the birth of the infant and may be presumed to seek the newborn's best interests. But they should be disqualified under conditions of child abuse, abandonment, neglect, and the like.

This suggested ranking should on occasion be reordered because of a decisionmaker's personal interest, ignorance, or bad faith. Serious conflicts of interest in the family may be more common than has generally been appreciated by either physicians or the courts. Many family members simultaneously have interests both in the patient's welfare and in the patient's death. A clear example is Case 5: Earle Spring's family was devoted to him but also was under a burdensome financial arrangement in paying for his care. As the debts mounted, a lien was placed on a family home. The court eventually appointed a guardian *ad litem* to investigate and protect Spring's interests. This case and many like it raise profound issues about valid familial decisionmaking under conflict of interest.

In many cases family members decline the decisionmaking role. Among the more difficult circumstances are those in which no party is available, willing,

UNIVERSIDAD RAFAEL LANDIVAR
BIBLIOTECA

or obligated to make decisions for incompetent persons. A health professional must either make the decision (with or without consultation) or wait until the patient's condition worsens so that an emergency can be declared or possibly seek judicial authorization or appointment of a guardian *ad litem*. Despite many distressing circumstances with few checks and little accountability, contemporary medicine has become accustomed to patients who have no surrogate and no hope for one. However, society has not yet come to grips with the problem. Below we consider a few approaches (ethics committees, ombudsmen, and the like) that should help alleviate this problem.

The role of health care professionals. Physicians and other health care professionals can help the family become adequate decisionmakers and can safeguard the patient's interests and preferences (where known) by monitoring the quality of surrogate decisionmaking. Physicians can sometimes discharge those obligations by withdrawing from the case or transferring the patient, but typically they have obligations to help patients and to ensure that surrogates do not violate those obligations. If a surrogate's decision is contested and differences cannot be resolved, the caregiver will need an independent source of review, a hospital ethics committee or the judicial system. In the event that a surrogate, a member of the health care team, or an independent reviewer asks a caregiver to perform an act that the caregiver regards as futile or unconscionable, the caregiver is not obligated to perform the act, but may still be obligated help the surrogate or patient make other arrangements for care.

In examining the role of physicians and other health care professionals, we need more empirical evidence about their willingness to override familial decisions and their reasons for doing so. Much of the available evidence is derived from parental decisions about neonates and is not accurate beyond this class of patients.[105] Some of this evidence indicates that physicians occasionally displace parents as decisionmakers to protect the parents rather than the infants.[106] Such paternalistic actions toward the parents of seriously ill newborns usually involve nondisclosure or manipulation of information rather than coercion. For example, physicians sometimes do not adequately inform parents on grounds that the information would overburden them, upset them, or make them feel guilty. These actions may be justified, but alternatives such as counseling also may alleviate the problems.

Institutional ethics committees. Surrogates or parents sometimes refuse treatments that are in the interests of those they should protect, and physicians sometimes too readily acquiesce. In other cases, decisionmakers need help in reaching difficult decisions. In both circumstances, a mechanism or procedure is needed to help make a decision or to break a closed, private circle of refusal

and acquiescence. A similar need exists for assistance in decisions regarding residents of nursing homes and hospices, psychiatric hospitals, and many residential facilities in which families often play no significant role. One promising, but loosely structured mechanism is the institutional ethics committee. Some state laws now mandate or legally empower these committees.[107]

Institutional committees were often established to allocate time on kidney machines, and for approximately three decades they have been required for research involving human subjects. However, their use for decisionmaking about treatment or nontreatment for incompetent patients is more recent and more controversial. According to a survey in 1981, ethics committees existed in only one percent of all hospitals, in less than five percent of the hospitals with more than two-hundred beds, and in no hospitals with fewer than two-hundred beds.[108] These committees grew dramatically in the 1980s to over sixty percent of all hospitals with two-hundred beds or more.[109] These committees differ widely in their composition and function. Many create and recommend explicit policies to govern actions such as withholding and withdrawing treatment, and many serve educational functions in the hospital. Controversy centers on additional functions such as whether, apart from evidence of abuse of incompetent patients, committees should make, facilitate, or monitor decisions about patients in particular cases.

Many people argue that informal de facto committees already exist because decisions about treatment and nontreatment are not private if numerous caregivers are involved, as is typical in institutions. Several cases have reached the courts because some member of the health care team, often a nurse, believed that a decision not to provide treatment violated legal obligations. From this perspective, committees are unnecessary in decisionmaking: They diffuse responsibility, impose another layer of bureaucratic delay if they are not convened in a timely way, and sometimes they become pawns of powerful groups.

The decisions of committees on occasion need to be reviewed or criticized, perhaps by an auditor or impartial fourth party. This procedural check is similar to the legal use of "neutral factfinders" appointed to monitor medical decisions made by parents for children who reject parental judgments.[110] Checks are needed because these committees are informal in their deliberations, and yet can have profound effects on families, institutions, and courts. The committees do not have formal procedures of evidence or legal representation, and checks help protect confidentiality, ensure fair representation, and provide for equal consideration.[111]

Nonetheless, the benefits of good committee review generally outweigh its risks. These committees help resolve disagreements, generate reasoned options, and help the parties conform to institutional guidelines and federal regulations. The committees also can help protect incompetent persons by facilitating treat-

ment when it has been unjustifiably refused, and by denying treatment when it should not have been authorized. Consider, for example, the following case, in which the committee was empowered to make a final decision:[112]

> [The review committee] found that the risks of the treatment outweighed its benefits [in the case of] an 85-year-old resident of a center for the developmentally disabled for whom a right inguinal hernia repair, cystoscopy, transurethral resection of the prostate, and right urethral-stone basketing were proposed. Citing the patient's frailty and the high risks of surgery for such a patient, the panel refused to consent to the procedures.

A major justification for committee review has been that open discussion and debate foster better thinking than can be expected of parties with a narrower perspective. The same justification is often given for clinical consultation. However, more research is needed on the role and functioning of these committees. We need to learn, for example, when committees satisfactorily serve as a forum for discussion without having a power of veto, when they should be empowered to make final decisions, and when they might engage in retrospective rather than prospective review of cases or in prospective review without veto power.

These committees have a particularly robust role to play in circumstances in which physicians acquiesce too readily to parental, familial, or guardian wishes. Until we better understand the extent to which families or guardians and physicians act or fail to act to pursue the best interests of infants, minors, or incompetent individuals, it is prudent and morally appropriate to require internal committee review whenever parents, families, or guardians decide that life-sustaining therapy should be forgone (whether or not the physician concurs with the surrogate's decision).[113] In some cases, it is appropriate to threaten parents with a possible court order to obtain necessary parental consent for a procedure that is clearly in a child's best interests.[114]

Surrogate review committees also can serve as a viable alternative to costly court review in the judicial system, especially when no clear legal conflict exists among the various parties to the decision. In some cases, committee advice can help the parties avoid threatened legal difficulties. However, these committees cannot be expected to settle serious legal disputes between parties, for reasons now to be discussed.

The judicial system. Courts have sometimes been unduly intrusive as final decisionmakers, but in many cases courts are the last recourse and the fairest decisionmaker. In a widely discussed declaration by a court, the supreme judicial court of Massachusetts held in *Saikewicz* that questions of life and death require the "process of detached but passionate investigation and decision that forms the ideal on which the judicial branch of government was created." This was expressly a departure from the *Quinlan* decision in New Jersey. The view from

Massachusetts was that the court has a responsibility for making these decisions that is "not to be entrusted to any other group." The court held that probate courts should make these decisions after considering all viewpoints and alternatives, including, if possible and appropriate, those of an ethics committee.[115]

In cases like *Saikewicz*, in which there is no involved family, another decisionmaker—physicians, a committee, or a court—is essential. Nothing is inherently objectionable about an appeal to probate courts as decisionmakers in such cases, but no solid evidence exists to indicate that physicians and hospital ethics committees would be less satisfactory than the courts in many cases. The courts should be invoked when there are good reasons to seek to disqualify the family or health care professionals in order to protect an incompetent patient's interests or to adjudicate conflicts over those interests. The courts also sometimes need to intervene in nontreatment decisions for salvageable incompetent patients in mental institutions, nursing homes, and the like. If no family members are available or willing to be involved, and if the patient is confined to a state mental institution or is in a nursing home, it may be appropriate to establish safeguards beyond the health care team and the institutional ethics committee. For example, the New Jersey Supreme Court in *Conroy* recommended the involvement of the state ombudsman, created several years earlier as an established administrative office for the surveillance of nursing homes.[116]

Conclusion

In this chapter we have concentrated on specifying the principle of nonmaleficence, particularly for actions that eventuate in death. Implicit throughout the chapter is the premise that morality is concerned with the harmfulness of harms *per se*, and not merely with responsibility for causing harm. If it is conceded that we can and should protect persons against some types and levels of harm, as well as avoid causing harm to them, it is a short step to the conclusion that a positive obligation exists to provide benefits such as health care. The step may be shorter still because of the conceptual and moral uncertainty that surrounds the distinctions between the obligation to avoid harm to others, the obligation to benefit them, and the obligation to treat them justly. These topics are engaged in Chapters 5 and 6.

Notes

1. W. H. S. Jones, *Hippocrates*, vol. I (Cambridge, MA: Harvard University Press, 1923), p. 165. See also Ludwig Edelstein, *Ancient Medicine*, ed. O. Temkin and C. L. Temkin (Baltimore: Johns Hopkins University Press, 1967); and Albert R.

Jonsen, "Do No Harm: Axiom of Medical Ethics," in *Philosophical and Medical Ethics: Its Nature and Significance,* ed. Stuart F. Spicker and H. Tristram Engelhardt, Jr. (Dordrecht, the Netherlands: D. Reidel, 1977), pp. 27–41.

2. See H. L. A. Hart, *The Concept of Law* (Oxford: Clarendon Press, 1961), p. 190.
3. See, for example, W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 21–26; and John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), p. 114.
4. William Frankena, *Ethics,* 2d Ed. (Englewood Cliffs, NJ: Prentice-Hall, 1973), p. 47.
5. *McFall v. Shimp,* no. 78–1771 in Equity (C. P. Allegheny County, Pa., July 26, 1978). See also Barbara J. Culliton, "Court Upholds Refusal to Be Medical Good Samaritan," *Science* 201 (August 18, 1978): 596–97; "Bone Marrow Transplant Plea Rejected," *American Medical News* 21 (August 11, 1978): 13; "Anemia Victim Dies, Asks Forgiveness for Cousin," *International Herald Tribune,* August 12–13, 1978; "Judge Upholds Transplant Denial," *New York Times,* July 27, 1978, p. A10; Dennis A. William and Lawrence Walsh, "The Law: Bad Samaritan," *Newsweek* 92 (August 7, 1978): 35.
6. Alan Meisel and Loren H. Roth, "Must a Man Be His Cousin's Keeper?" *Hastings Center Report* 8 (October 1978): 5–6.
7. W. D. Ross, for example, regards "not injuring others" as a synonym of "nonmaleficence" and includes under the duty of nonmaleficence several of the Decalogue's prohibitions of harmful actions, such as killing, stealing, committing adultery, and bearing false witness. Ross, *The Right and the Good,* pp. 21–22.
8. See Joel Feinberg, *Harm to Others,* vol. I of *The Moral Limits of the Criminal Law* (New York: Oxford University Press, 1984), pp. 32–36.
9. For rules of nonmaleficence, see Bernard Gert, *Morality: A New Justification of Morality* (New York: Oxford University Press, 1988), ch. 6–7. He offers a different account of justificatory support for moral rules.
10. For a careful criticism of the priority of avoiding harm, see Nancy Davis, "The Priority of Avoiding Harm," in *Killing and Letting Die,* ed. Bonnie Steinbock (Englewood Cliffs, NJ: Prentice-Hall, 1980), pp. 172–214.
11. See Eric D'Arcy, *Human Acts: An Essay in their Moral Evaluation* (Oxford: Clarendon Press, 1963), p. 121.
12. Cf. William L. Prosser, *Handbook of the Law of Torts,* 4th Ed. (St. Paul, MN: West Publishing, 1971), pp. 145–46. For a broad view that includes "moral negligence," see Ronald D. Milo, *Immorality* (Princeton, NJ: Princeton University Press, 1984).
13. See "Physician's Duty to Inform of Risks," *American Law Reports,* 3d, 88 (1986): 1010–25; and Martin Curd and Larry May, *Professional Responsibility for Harmful Actions* (Dubuque, IA: Kendall/Hunt, 1984).
14. Quoted in Angela Roddy Holder, *Medical Malpractice Law* (New York: John Wiley & Sons, 1975), p. 42.
15. This case was presented to one of the authors during a consultation.
16. *In the matter of Spring,* Mass. 405 N.E. 2d 115 (1980), at 488–89.
17. See President's Commission, *Deciding to Forego Life-Sustaining Treatment,* pp. 73–77.
18. Robert Stinson and Peggy Stinson, *The Long Dying of Baby Andrew* (Boston: Little, Brown and Co., 1983), p. 355.
19. Susanna E. Bedell and Thomas L. Delbanco, "Choices about Cardiopulmonary Resuscitation in the Hospital: When Do Physicians Talk with Patients?," *New England Journal of Medicine* 310 (April 26, 1984): 1089–93. See also Marcia Angell, "Respecting the Autonomy of Competent Patients," *New England Journal of Medicine* 310 (April 26, 1984): 1115–16.
20. See Paul Ramsey, *The Patient as Person* (New Haven: Yale University Press, 1970), p. 120.
21. See Diane Lynn Redleaf, Suzanne Baillie Schmitt, and William Charles Thompson, "The California Natural Death Act: An Empirical Study of Physicians' Practices," *Stanford Law Review* 31 (May 1979): 913–47.
22. Gerald Kelly, S.J., "The Duty to Preserve Life," *Theological Studies* 12 (December 1951): 550.
23. This case has been adapted with permission from a case presented by Dr. Martin P. Albert of Charlottesville, VA.
24. In the matter of *Quinlan,* 70 N.J. 10, 355 A.2d 647, *cert. denied,* 429 U.S. 922 (1976).
25. In re *Conroy,* 486 A.2d 1209 (N.J. 1985).
26. *Brophy v. New England Sinai Hospital, Inc.,* 398 Mass. 417, 497 N.E. 2d 626 (1986).
27. These issues were first raised in 1982, in *Barber v. Superior Court,* 147 Cal. App. 3d 1006, 195 Cal. Rptr. 484 (1983). By 1988, many courts accepted this trend as determinative [see, e.g., *Gray v. Romeo,* 697 F.Supp. 580 (D.R.I. 1988) and *McConnell v. Beverly Enterprises,* 209 Conn. 692 (Conn. Sup. Ct. 1989), 553 A.2d 596]. For a review of the massive court literature during this formative period, see Alan Meisel, *The Right to Die* (New York: John Wiley and Sons, 1989), § 5.10. In *Cruzan v. Director, Missouri Department of Health,* 110 S.Ct. 2841 (1990), the U.S. Supreme Court focused on procedural requirements for termination of life-sustaining treatment for incompetent patients. The court assumed that a competent person has a constitutionally protected right to refuse lifesaving hydration and nutrition. Its dicta reflected no distinction between medical and sustenance treatments.
28. See Joanne Lynn and James F. Childress, "Must Patients Always Be Given Food and Water?" *Hastings Center Report* 13 (October 1983): 17–21. See also the essays in Joanne Lynn, ed., *By No Extraordinary Means* (Bloomington: Indiana University Press, 1986).
29. G. E. M. Anscombe, "Ethical Problems in the Management of Some Severely Handicapped Children: Commentary," *Journal of Medical Ethics* 7 (1981): 122.
30. Koop, "Ethical and Surgical Considerations in the Care of the Newborn with Congenital Abnormalities," in *Infanticide and the Handicapped Newborn,* ed. Dennis J. Horan and Melinda Delahoyde (Provo, UT: Brigham Young University Press, 1982), pp. 89–106, esp. 105.
31. C. Everett Koop and Edward R. Grant, "The 'Small Beginnings' of Euthanasia," *Journal of Law, Ethics & Public Policy* 2 (1986): 607–32.
32. *Federal Register* 48, No. 129, July 5, 1983.
33. See the summary of living will legislation in *The Physician and the Hopelessly Ill Patient* (New York: Society for the Right to Die, 1985), pp. 39–80, and *1988 Supplement,* pp. 17–34.
34. Daniel Callahan, "On Feeding the Dying," *Hastings Center Report* 13 (October 1983): 22, and see Ronald A. Carson, "The Symbolic Significance of Giving to Eat and Drink," in *By No Extraordinary Means,* pp. 85, 87.

35. See Mark Siegler and Alan J. Weisbard, "Against the Emerging Stream: Should Fluids and Nutritional Support Be Discontinued?" *Archives of Internal Medicine* 145 (January 1985): 129–32; and Patrick Derr, "Why Food and Fluids Can Never Be Denied," *Hastings Center Report* 16 (February 1986): 28–30.
36. Joyce V. Zerwekh, "The Dehydration Question," *Nursing '83* (January 1983): 47–51, reprinted in Lynn, ed., *By No Extraordinary Means*, ch. 2; Ronald Cranford, "Neurologic Syndromes and Prolonged Survival: When Can Artificial Nutrition and Hydration be Foregone?" *Law, Medicine, and Health Care* 19 (1991): 13–22, esp. 18–19.
37. Kenneth C. Micetich, Patricia H. Steinecker, and David C. Thomasma, "Are Intravenous Fluids Morally Required for Dying Patients?" *Archives of Internal Medicine* 143 (May, 1983): 975–78.
38. President's Commission, *Deciding to Forego Life-Sustaining Treatment*, pp. 17–18.
39. The rule of double effect has precedents that predate the writings of St. Thomas Aquinas (e.g., in Augustine and Abelard). However, the history primarily flows from Aquinas in traditions such as that of the Jesuits. See Anthony Kenny, "The History of Intention in Ethics," *Anatomy of the Soul* (Oxford: Basil Blackwell, 1973), Appendix; and Joseph T. Mangan, S.J., "An Historical Analysis of the Principle of Double Effect," *Theological Studies* 10 (1949): 41–61.
40. However, the RDE is defended by some as having only a prima facie moral force that can be overridden by competing moral considerations. See Warren S. Quinn, "Actions, Intentions, and Consequences: The Doctrine of Double Effect," *Philosophy and Public Affairs* 18 (1989): 334–51, esp. 344–45.
41. Joseph Boyle reduces the RDE to two conditions: intention and proportionality. "Who Is Entitled to Double Effect?" *Journal of Medicine and Philosophy* 16 (1991): 475–94, and "Toward Understanding the Principle of Double Effect," *Ethics* 90 (1980): 527–38. For an emphasis on intention, see Charles Fried, *Right and Wrong* (Cambridge, MA: Harvard University Press, 1978) and Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986). For an emphasis on proportionality, see Richard McCormick, *Ambiguity in Moral Choice* (Milwaukee, WI: Marquette University, 1973) and his contribution to Paul Ramsey and Richard A. McCormick, S.J., eds., *Doing Evil to Achieve Good: Moral Choice in Conflict Situations* (Chicago: Loyola University Press, 1978). For guides to the large literature, see works cited in this section and several essays in the *Journal of Medicine and Philosophy* 16 (1991).
42. For these cases, see David Granfield, *The Abortion Decision* (Garden City, NY: Image Books, 1971), which defends the RDE, and Susan Nicholson, *Abortion and the Roman Catholic Church* (Knoxville, TN: Religious Ethics, Inc., 1978), which criticizes it. See also the criticisms in Donald Marquis, "Four Versions of Double Effect," *Journal of Medicine and Philosophy* 16 (1991): 515–44.
43. For a critique with special reference to Richard McCormick, see G. E. M. Anscombe, "Action, Intention, and 'Double Effect,' " *Proceedings of the American Catholic Philosophical Association* 56 (1982): 21–24.
44. The most developed analysis is Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).
45. Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970), pp. 49–85.

46. See Hector-Neri Castañeda, "Intensionality and Identity in Human Action and Philosophical Method," *Nous* 13 (1979): 235–60, esp. 255.
47. Our analysis here borrows from Ruth R. Faden and Tom L. Beauchamp, *A History and Theory of Informed Consent* (New York: Oxford University Press, 1986), chap. 7.
48. We follow John Searle in thinking that we cannot reliably distinguish in many situations among acts, effects, consequences, and events. Searle, "The Intentionality of Intention and Action," *Cognitive Science* 4 (1980): 65.
49. Such an interpretation of double effect is defended by Boyle, "Who Is Entitled to Double Effect?"
50. See the synoptic argument to this conclusion in Helga Kuhse, *The Sanctity-of-Life Doctrine in Medicine: A Critique* (Oxford: Clarendon Press, 1987), esp. pp. 93–103.
51. For debates about futility, see Stuart J. Youngner, "Who Defines Futility?," *Journal of the American Medical Association* 260 (October 14, 1988): 2094–95 and "Futility in Context," *Journal of the American Medical Association* 264 (September 12, 1990): 1295–96; Steven H. Miles, "Informed Demand for 'Non-Beneficial' Medical Treatment," *New England Journal of Medicine* 325 (Aug. 15, 1991): 512–15; John D. Lantos et al., "The Illusion of Futility in Clinical Practice," *The American Journal of Medicine* 87 (July 1989): 81–84; and Nancy Jecker, "Are Physicians Obligated to Provide Futile Treatment?," *Medical Ethics* 7 (December 1992): 9–11.
52. This case was recorded by Robert Baker in his project on Moral Methodologies in ICUs.
53. Many judicial opinions accept this conclusion. Competent patients have the legal right to refuse treatment in the United States, even if not terminally ill, as a result of *Cruzan v. Director*, 110 S.Ct. 2841, at 2851 (1990). See also *Bouvia* and *McAfee*, as discussed later in this chapter. For statements regarding incompetent patients, see *In re Browning*, 543 So.2d 258 (Fla. Dist. Ct. App. 1989), *aff'd* 568 So.2d 4 (Fla. 1990) and *McConnell v. Beverly Enterprises*, 209 Conn. 692, 553 A.2d 596 (1989).
54. *Superintendent of Belchertown State School v. Saikewicz*, Mass., 370 N.E. 2d 417 (1977), at 428.
55. Ibid.
56. Ramsey, *Ethics at the Edges of Life* (New Haven: Yale University Press, 1978), p. 155.
57. See President's Commission, *Deciding to Forego Life-Sustaining Treatment*, ch. 5, and the articles on "The Persistent Problem of PVS," *Hastings Center Report* 18 (February/March 1988): 26–47.
58. Ramsey, *Ethics at the Edges of Life*, p. 172.
59. President's Commission, *Deciding to Forego Life-Sustaining Treatment*.
60. See Norman L. Cantor, *Legal Frontiers of Death and Dying* (Bloomington: Indiana University Press, 1987), pp. 87–91.
61. See Frank A. Chervenak and Laurence B. McCullough, "Nonaggressive Obstetric Management," *Journal of the American Medical Association* 261 (June 16, 1989): 3439–40. For issues about decisions regarding seriously ill newborns, see Hastings Center Newborn Project, "Imperiled Newborns," *Hastings Center Report* 17 (December 1987): 5–32; Richard C. McMillan, H. Tristram Engelhardt, Jr., and Stuart

F. Spicker, eds., *Euthanasia and the Newborn: Conflicts Regarding Saving Lives* (Dordrecht, the Netherlands: D. Reidel, 1987); and Arthur L. Caplan and Robert H. Blank, eds., *Compassion: Government Intervention in the Treatment of Critically Ill Newborns* (Totowa, NJ: The Humana Press, 1992).

62. Albert R. Jonsen and Michael J. Garland, "A Moral Policy for Life/Death Decisions in the Intensive Care Nursery," in *Ethics of Newborn Intensive Care*, ed. Albert R. Jonsen and Michael J. Garland (Berkeley: University of California, Institute of Governmental Studies, 1976), p. 148.
63. Raymond S. Duff and A. G. M. Campbell, "Moral and Ethical Dilemmas in the Special-Care Nursery," *New England Journal of Medicine* 289 (October 25, 1973): 890–94.
64. "Child Abuse Prevention and Treatment and Adoption Reform Act Amendments of 1984," Public Law 98–457, 42 U.S.C. 5101ff (1984); "Child Abuse and Neglect Prevention and Treatment Program: Final Rule," *Federal Register* 50 (April 15, 1985): 14878–901. The American Medical Association, the American Hospital Association and other professional groups challenged these regulations in *United States v. University Hospital*, 729 F.2d 144 (1984) and *American Hospital Association v. Heckler*, 585 F. Supp. 541, App. to Pet. for Cert. 50a (1984). Lower courts found the regulations invalid, and this ruling was upheld by the U.S. Supreme Court on June 9, 1986 (*Bowen v. American Hospital Association et al.*, No. 84-1529, 54 LW 4579 (June 9, 1986). For stages in the evolution of the government's action, see Nancy M. P. King, "Federal and State Regulation of Neonatal Decision-Making," and Mary Ann Gardell and H. Tristram Engelhardt, Jr., "The Baby Doe Controversy: An Outline of Some Points in Its Development," in McMillan, Engelhardt, and Spicker, eds., *Euthanasia and the Newborn*, pp. 89–116, 293–99.
65. See Helga Kuhse and Peter Singer, *Should the Baby Live?* (Oxford: Oxford University Press, 1985), p. 46.
66. See Fred Barbash and Christina Russell, "The Demise of 'Infant . . .': Permitted Death Gives Life to an Old Debate," *Washington Post*, April 17, 1982.
67. *In re Colyer*, 660 P.2d 738 (1983), at 751.
68. A special issue of *Law, Medicine, and Health Care* 17 (Winter 1989) is devoted to the Linares case. See also Steven H. Miles, "Taking Hostages: The Linares Case," *Hastings Center Report* 19 (July/August 1989): 4.
69. In the matter of *Claire C. Conroy*, 190 N.J. Sup. 453, 464 A.2d 303 (App. Div. 1983). See the analysis in several articles in Joanne Lynn, ed., *By No Extraordinary Means*, pp. 227–66.
70. In the matter of *Claire C. Conroy*, 486 A.2d 1209 (New Jersey Supreme Court, 1985), at 1222–23, 1236.
71. See *In re Estate of Greenspan*, 558 N.E.2d 1194, at 1203 (Ill. 1990).
72. Causal judgments are *commonly* controlled by forms of practical judgment as well as by causal explanation. See H. L. A. Hart and A. M. Honore, *Causation in the Law* (Oxford: Clarendon Press, 1959), and Samuel Gorovitz, "Causal Judgments and Causal Explanations," *Journal of Philosophy* 62 (1965): 695–711.
73. For an example of one author who straightforwardly uses "killing" as a moral notion entailing culpability, see Daniel Callahan, "When Self-Determination Runs Amok," *Hastings Center Report* 22 (March–April 1992): 53–54.
74. Dan Brock, "Death and Dying," in *Medical Ethics*, ed. Robert M. Veatch (Boston: Jones and Bartlett Publishers, 1989), p. 345.

75. American Geriatrics Society, Public Policy Committee, "Voluntary Active Euthanasia," *Journal of the American Geriatrics Society* 39 (August 1991): 826.
76. American Medical Association, Council on Ethical and Judicial Affairs, *Euthanasia: Report C*, in *Proceedings of the House of Delegates* (Chicago: American Medical Association, June, 1988): 258–60 (and see *Current Opinions*, § 2.20, p. 13, 1989); "Decisions Near the End of Life," *Report B*, adopted by the House of Delegates (1991), pp. 11–15, and see the abridged version in "Decisions Near the End of Life," *Journal of the American Medical Association* 267 (April 22/29, 1992): 2229–33. In March 1986 the AMA amended an earlier statement so that "life-prolonging medical treatment includes medication and artificially or technologically supplied respiration, nutrition or hydration." *Current Opinions* (Chicago: AMA, 1986), § 2.18, pp. 12–13.
77. *People v. Roberts*, 211 Mich. 187, 178 N.W. 690 (1920).
78. This distinction and our arguments are indebted to John Rawls, "Two Concepts of Rules," *Philosophical Review* 64 (1955): 3–32.
79. For fuller discussions, see Douglas Walton, *Slippery Slope Arguments* (Oxford: Clarendon Press, 1992); Trudy Govier, "What's Wrong with Slippery Slope Arguments?" *Canadian Journal of Philosophy* 12 (June 1982): 303–16; Frederick Schauer, "Slippery Slopes," *Harvard Law Review* 99 (1985): 361–83; Bernard Williams, "Which Slopes Are Slippery?" in *Moral Dilemmas in Modern Medicine*, ed. Michael Lockwood (Oxford: Oxford University Press, 1985), pp. 126–37; David Lamb, *Down the Slippery Slope: Arguing in Applied Ethics* (London: Croom Helm, 1988); Wibren van der Burg, "The Slippery Slope Arguments," *Ethics* 102 (October 1991): 42–65; and James Rachels, *The End of Life: Euthanasia and Morality* (Oxford: Oxford University Press, 1986), ch. 10.
80. See Gerald J. Hughes, S.J., "Killing and Letting Die," *The Month* 8 (February 1975): 42–45; and David Louisell, "Euthanasia and Biathanasia: On Dying and Killing," *Linacre Quarterly* 40 (November 1973): 234–58.
81. Leo Alexander, "Medical Science under Dictatorship," *New England Journal of Medicine* 241 (1949): 39–47.
82. Robert Jay Lifton, *The Nazi Doctors: Medical Killing and the Psychology of Genocide* (New York: Basic Books, 1986).
83. See Rachels, *The End of Life*, and the special supplement on "Biomedical Ethics and the Shadow of Nazism," *Hastings Center Report* 6 (August 1976), Supp., esp. the article by Lucy Dawidowicz. See also Arthur L. Caplan, ed., *When Medicine Went Mad: Bioethics and the Holocaust* (Totowa, NJ: Humana Press, 1992); and George J. Annas and Michael Grodin, *The Nazi Doctors and the Nuremberg Code* (New York: Oxford University Press, 1992).
84. "It's Over, Debbie," *Journal of the American Medical Association* 259 (1988): 272. A substantial controversy followed publication of this article. The pro and con positions are quoted and much of the literature summarized in Victor Cohn, "Is It Time for Mercy Killing?," *Washington Post*, Health Section (Aug. 15, 1989), pp. 12–15.
85. We owe much in this argument to James Rachels (personal correspondence).
86. Marcia Angell, "The Quality of Mercy," *New England Journal of Medicine* 306 (January 14, 1982): 98–99.
87. Paul J. van der Maas et al., "Euthanasia and Other Medical Decisions Concerning the End of Life," *The Lancet* 338 (September 14, 1991): 669–74. Defenders and critics of euthanasia interpret the evidence differently. For a critic's interpretation,

see Carlos Gomez, *Regulating Death: The Case of the Netherlands* (New York: The Free Press, 1991); for a defender's interpretation, see Margaret Battin, "Voluntary Euthanasia and the Risks of Abuse: Can We Learn Anything from the Netherlands?" *Law, Medicine & Health Care* 20 (Spring–Summer, 1992): 135.

88. For a discussion of this case, see Paige Mitchell, *Act of Love: The Killing of George Zygmaniak* (New York: Alfred A. Knopf, 1976).
89. Based on: *New York Times,* June 6, pp. A1, B6; June 7, 1990, pp. A1, D22; June 9, p. A6; June 12, p. C3; *Newsweek,* June 18, 1990, p. 46. For Kevorkian's description of the events, see his *Prescription: Medicide* (Buffalo, NY: Prometheus Books, 1991), pp. 221–31.
90. *State of Georgia v. McAfee,* 385 S.E.2d 651 (Ga. 1989).
91. *Bouvia v. Superior Court,* 179 Cal. App. 3d 1127, at 1146–47. 225 Cal. Rptr. 297 (1986). See also Lorna A. Voboril, "*Bouvia v. Superior Court:* The Death Option," *Pacific Law Journal* 18 (1987): 1029–53.
92. See Timothy E. Quill, "Death and Dignity: A Case of Individualized Decision Making," *New England Journal of Medicine* 324 (March 7, 1991): 691–94, reprinted with additional analysis in Quill, *Death and Dignity* (New York: W. W. Norton & Co., 1993); Lawrence K. Altman, "A Doctor Agonized but Provided Drugs To Help End a Life," *New York Times,* March 3, 1991, pp. A1, B8; Altman, "Jury Declines to Indict a Doctor Who Said He Aided in a Suicide," *New York Times,* July 27, 1991, pp. A1, A10.
93. For a recommendation that we should emphasize good palliative care and sensitive laws for termination of treatment as our public policy, while rejecting active killing, see Joanne Lynn, "The Health Care Professional's Role When Active Euthanasia Is Sought," *Journal of Palliative Care* 4 (1988): 100–102, and Susan Wolf, "Holding the Line on Euthanasia," *Hastings Center Report* 19 (January–February 1989): S13–S15.
94. See Rawls, *A Theory of Justice,* pp. 85–86.
95. In 1991, forty-nine states and the District of Columbia had some form of advance-directive law that anticipates decisions and decisionmaking authority regarding use of life-sustaining treatment. Of these, twenty states acknowledged durable power of attorney permitting the appointment of a surrogate to make treatment decisions; this figure showed a dramatic increase, since only nine acknowledged durable power of attorney in 1989. See Advance Directives Seminar Group, "Advance Directives: Are They an Advance?," *Canadian Medical Association Journal* 146 (January 15, 1992): 127–34. For a helpful discussion, see Nancy M. P. King, *Making Sense of Advance Directives* (Dordrecht, the Netherlands: Kluwer Academic Publishers, 1991).
96. For an attempt to combine the two forms in a single document, see Robert Olick, "Approximating Informed Consent and Fostering Communication: The Anatomy of an Advance Directive," *Journal of Clinical Ethics* 2 (1991): 181–95.
97. For a balanced account of problems and promise in advance directives, see Dan Brock, "Trumping Advance Directives," *Hastings Center Report* 21 (September–October 1991): S5–S6.
98. See E. R. Gamble et al., "Knowledge, Attitudes, and Behavior of Elderly Persons Regarding Living Wills," *Archives of Internal Medicine* 151 (February 1991): 277–80.
99. General Assembly of Pennsylvania, Senate Bill No. 3, Session of 1991, as amended April 6, 1992, as published in *Philadelphia Medicine* 88 (August 1992): 329–33.

100. Stuart J. Eisendrath and Albert R. Jonsen, "The Living Will," *Journal of the American Medical Association* 249 (April 15, 1983): 2054–58.
101. See the somewhat different set of qualifications proposed by Robert Weir, *Selective Nontreatment of Handicapped Newborns* (New York: Oxford University Press, 1984), ch. 9. He uses the criterion of impartiality, rather than partiality.
102. See Judith Areen, "The Legal Status of Consent Obtained from Families of Adult Patients to Withhold or Withdraw Treatment," *Journal of the American Medical Association* 258 (July 10, 1987): 229–35; and John W. Warren et al., "Informed Consent by Proxy: An Issue in Research with Elderly Patients," *New England Journal of Medicine* 315 (October, 1986): 1124–28, as discussed in Chapter 3, this volume (p. 177).
103. Nancy Rhoden has proposed that physicians who reject family's choices should be placed under a burden of adducing evidence in court as to the unreasonableness of the decision. "Litigating Life and Death," *Harvard Law Review* 102 (1988): 437. Patricia King has argued that some family interests other than those that support the patient's best medical interests should be considered legitimate. "The Authority of Families to Make Medical Decisions for Incompetent Patients after the *Cruzan* Decision," *Law, Medicine & Health Care* 19 (1991): 76–79.
104. Virginia Natural Death Act, Va. Code §§ 54–325.8:1–13 (1983).
105. See A. Shaw, J. G. Randolph, and B. Manard, "Ethical Issues in Pediatric Surgery: A National Survey of Pediatricians and Pediatric Surgeons," *Pediatrics* 60 (1977): 588–99; and David Todres, "Pediatricians' Attitudes Affecting Decision-Making in Defective Infants," *Pediatrics* 60 (1977): 197. Even though there is debate about the available survey data, such data played a role in the efforts by the U.S. Department of Health and Human Services to establish a mechanism for protecting handicapped infants from discrimination. See *Federal Register* 49 (January 12, 1984): 1645.
106. See President's Commission, *Deciding to Forego Life-Sustaining Treatment,* pp. 210–11.
107. See Clarence J. Sundram, "Informed Consent for Major Medical Treatment of Mentally Disabled People," *New England Journal of Medicine* 318 (May 26, 1988): 1368–73.
108. President's Commission, *Deciding to Forego Life-Sustaining Treatment,* p. 446.
109. See American Academy of Pediatrics Infant Bioethics Task Force and Consultants, Guidelines for Infant Bioethics Committees, *Pediatrics* 74 (August 1984): 306–310; American Medical Association, Judicial Council, "Guidelines for Ethics Committees in Health Care Institutions," *Journal of the American Medical Association* 253 (May 10, 1985): 2698–99; Fred Rosner, "Hospital Medical Ethics Committees: A Review of Their Development," *Journal of the American Medical Association* 253 (May 10, 1985); Tracy Miller and Anna Maria Cugliari, "Withdrawing and Withholding Treatment: Policies in Long-Term Care Facilities," *Gerontologist* 30 (August 1990): 462–68.
110. See *Parham v. J.R.,* 442 U.S. 584, 602 (1979).
111. See Susan M. Wolf, "Ethics Committees and Due Process: Nesting Rights in a Community of Caring," *Maryland Law Review* 50 (1991): 798–858.
112. Sundram, "Informed Consent for Major Medical Treatment of Mentally Disabled People," p. 1372.
113. Cf. President's Commission, *Deciding to Forego Life-Sustaining Treatment,* p. 227. Contrast Raymond S. Duff and A. G. M. Campbell, "Moral Communities

and Tragic Choice," in McMillan, Engelhardt, and Spicker, eds., *Euthanasia and the Newborn*, pp. 273–80.

114. For a case in which this expedient was necessary, see Mary B. Mahowald, "Baby Doe Committees: A Critical Evaluation," Ethical and Legal Issues in Perinatology, *Clinics in Perinatology* 15 (December 1988): 792–93.
115. *Superintendent of Belchertown State School v. Saikewicz*, Mass., 370 N.E. 2d 417 (1977).
116. *In re Conroy*, 486 A.2d 1209 (N.J. 1985), at 1239–42.

5

Beneficence

Morality requires not only that we treat persons autonomously and refrain from harming them, but also that we contribute to their welfare. Such beneficial actions fall under the heading of beneficence. No sharp breaks exist on the continuum from the noninfliction of harm to the provision of benefit, but principles of beneficence potentially demand more than the principle of nonmaleficence because agents must take positive steps to help others, not merely refrain from harmful acts. The word *nonmaleficence* is sometimes used broadly to include the prevention of harm and the removal of harmful conditions. However, prevention and removal require positive acts to benefit others, and therefore belong under beneficence rather than nonmaleficence.

In the present chapter we examine two principles of beneficence: positive beneficence and utility. *Positive beneficence* requires the provision of benefits. *Utility* requires that benefits and drawbacks be balanced. Both together are distinguished from the virtue of benevolence, from various forms of care, and from nonobligatory ideals of beneficence. Building on these distinctions and our analysis of them, we discuss conflicts between beneficence and respect for autonomy in paternalistic refusals to acquiesce in a patient's wishes or choices. The remainder of the chapter focuses on balancing benefits, risks, and costs, especially through analytical methods designed to implement the principle of utility in health policy and clinical care. We conclude that these methods have a useful, although limited, role as aids to decisionmaking, and that they need