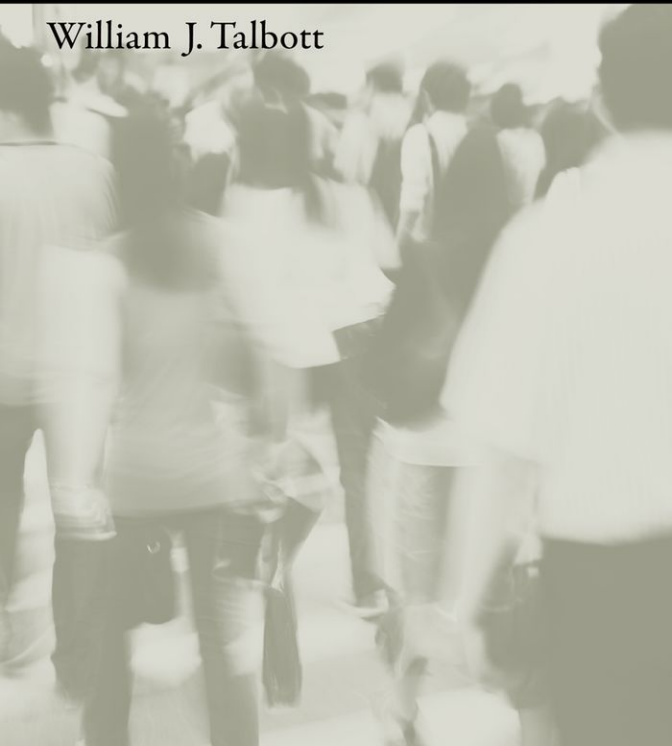# Human Rights
and Human Well-Being

William J. Talbott

HUMAN RIGHTS AND HUMAN WELL-BEING

OXFORD POLITICAL PHILOSOPHY

GENERAL EDITOR: SAMUEL FREEMAN,
UNIVERSITY OF PENNSYLVANIA

Oxford Political Philosophy publishes books on theoretical and applied political philosophy within the Anglo-American tradition. The series welcomes submissions on social, political, and global justice, individual rights, democracy, liberalism, socialism, and constitutionalism.

N. Scott Arnold
Imposing Values: An Essay on Liberalism and Regulation

Peter de Marneffe
Liberalism and Prostitution

William J. Talbott
Human Rights and Human Well-Being

HUMAN RIGHTS AND HUMAN WELL-BEING

William J. Talbott

OXFORD
UNIVERSITY PRESS

In Memory of Robert Nozick,
Teacher and Friend

This page intentionally left blank

I have the innate duty . . . so to affect posterity through each member of the sequence of generations in which I live, simply as a human being, that future generations will become continually better . . . and that this duty may thus rightfully be passed on from one generation to the next. . . . Without this hope for better times the human heart would never have been warmed by a serious desire to do something useful for the common good.

—Immanuel Kant

This page intentionally left blank

Acknowledgments

This is the second of two volumes on human rights that attempt to answer the question "Which rights should be universal?" and to explain why they should be. I have already thanked those who helped with the first volume. I have many more people to thank for helping me to complete this project.

Let me begin by thanking my students. I was fortunate to be able to teach an early draft of this book in a seminar at the University of Washington in spring 2005. I am grateful to the students in that seminar for much valuable feedback: Ben Almassi, Daniel Baker, Ron Belgau, Erika Dahlstrom, Ethel Evans-Elison, Jeremy Fischer, Leann Haggard, Philip McGrane, Scott McMahon, Brandon Morgan-Olsen, Kelly O'Connell, Dustin Pearson, Gabriela Remow, Benjamin Robbins, Christi Siver, and Edward Wolcher. I am also grateful to students in my course on liberty for feedback on my consequentialist account of rights against paternalism, especially to Adam Caldwell, John Gresham, Quinn Rotchford, and Shawn Stevenson.

One advantage of writing two volumes on human rights is that I have been able to take account of responses to the first volume in the second. I have been fortunate to have received many cogent criticisms of the first volume. At an APA Pacific Division session in March 2006, my critics included Carol Gould, James Nickel, and David Reidy. Gould and Reidy's criticisms and an additional critical review by Jeppe von Platz were published with my replies (Talbott 2008). At a symposium at the University of Washington in April 2006, I was fortunate to have Henry Shue and Kok-Chor Tan as my critics. Henry Shue reprised his role at an APSA session in August 2006, where he led a formidable lineup of critics that included Brooke Ackerly, Charles Beitz, and Jack Donnelly. Their criticisms and my replies were subsequently published (Talbott 2007). Finally, I received criticisms from Christopher Knapp at an APA Eastern Division session in December 2007. At all of these sessions, I also received valuable feedback from members of the audience.

In addition to opportunities to respond to criticisms of the first volume, I was fortunate to receive a number of invitations that allowed me to try out some parts of this second volume before publishing it. I presented an earlier version of the first three chapters at the NYU Colloquium in Legal, Social, and Political Philosophy in September 2006. I am especially grateful to Jeremy Waldron for his sympathetic presentation of my view, to Ronald Dworkin and

This book really began in an undergraduate political philosophy class taught by T. M. Scanlon in the spring of 1968. It was in that course that I read J. S. Mill's *On Liberty* and a draft of Scanlon's "A Theory of Freedom of Expression" for the first time. Although Scanlon was not a consequentialist,

both he and Mill provided models of indirect arguments for autonomy rights—that is, arguments for autonomy rights not grounded in the intrinsic value of autonomy. The following year I did an independent study with Scanlon on the manuscript of John Rawls's *Theory of Justice*. Again I was struck by Rawls's attempt to ground the autonomy rights of his special conception of justice indirectly in what seemed to me to be the well-being considerations of his general conception. This work with Scanlon was formative for my philosophical outlook, and I am deeply appreciative of his contribution to my thinking.

I first began to articulate my own attempt to ground individual rights indirectly on considerations of well-being when, as a graduate student, I took Robert Nozick's course in political philosophy in the fall of 1972. In that course, Nozick taught the manuscript of his philosophical defense of libertarianism, *Anarchy, State, and Utopia*. I wrote a term paper in which I criticized Nozick's libertarian position from what, in retrospect, I can see was a kind of Millian-Rawlsian indirect consequentialist viewpoint. The paper was a distant precursor of the indirect consequentialist position that I defend in this book. It led to lots of good philosophical give-and-take. Anyone who knew Nozick can tell you that I could not have had a more trenchant critic. He was also a warm human being and a source of unstinting encouragement.

I always knew that someday I would revise that term paper. Finally, 37 years later, I am ready to turn in the final draft. Sadly, my professor is no longer able to accept late work.

This page intentionally left blank

Contents

This page intentionally left blank

HUMAN RIGHTS AND HUMAN WELL-BEING

This page intentionally left blank

The Consequentialist Project
for Human Rights

In this, the second of two volumes, I continue the project of explaining which rights should be universally guaranteed to all normal human adults by governments everywhere.[1] In the first volume I focused on what I regard as the basic human rights. In this book I discuss both basic and nonbasic human rights and explain more fully why the rights I discuss, both basic and nonbasic, should be universal. I have written this book to stand on its own, so that it is not necessary to have read the first volume before reading this one.

My goal is to contribute to an important explanatory project in political philosophy. In this chapter I say what the project is and provide an overview of how I propose to contribute to it. The first volume dealt extensively with the metaphysics and the epistemology of moral belief. In this chapter, I review that discussion briefly and then, in chapters 7 and 8, I develop the epistemology more fully.

Mill's and Rawls's Consequentialist Projects

Perhaps the best way to introduce the project of this book is to do so historically. The project began in the 1850s with J. S. Mill's *On Liberty* [1859].[2] Mill's book was to be a new kind of defense of a package of autonomy rights, including rights to freedom of thought and discussion, freedom of the press, freedom of association, and freedom from paternalism. Mill was not the first philosopher to defend a package of autonomy rights. Locke and Kant, among many others, had defended such rights long before Mill. What made Mill's defense of them distinctive was that he did not begin by assuming such rights or by assuming that they were to be justified by the intrinsic value of autonomy. He intended to show that a package of autonomy rights could be justified on utilitarian grounds—that is, on the basis of the contribution to overall well-being that would result from the government's legally enforcing them.

Would the rights be absolute, so that no exceptions could ever be justified? Though Mill sometimes wrote in a way that suggested the rights should be absolute, from his discussion of examples it is clear that he did allow for exceptions.[3] This is not surprising, because even most nonconsequentialists allow that rights have some exceptions.[4]

At the time that Mill was writing, it was generally assumed that the only kind of rights that a utilitarian could justify were rights the government should infringe whenever it thought that infringing them would maximize overall utility. Call this sort of right an *act utilitarian* right. Mill was aiming to defend rights much stronger than this. Even if autonomy rights should not be absolute, Mill would argue on utilitarian grounds that autonomy rights ought to be stronger than act utilitarian rights—strong enough, that is, that a government could not justify infringing them simply because the government thought that the infringement would maximize overall utility. Indeed, because Mill regarded autonomy rights as protections against not only government tyranny, but also tyranny by a majority, he clearly intended that they be strong enough not to be overridden by a simple majority. I refer to rights of this kind as *robust rights*. Because robust rights need not be absolute, there is no presumption that they can never be overridden, only that what is necessary to override them is significantly more demanding than what is necessary to override act utilitarian rights. Thus, for example, there is no paradox in thinking that in enforcing such rights the courts would sometimes have to prohibit actions that the government believes will maximize overall utility or to invalidate legislation adopted by a majority vote in the legislature.

There was one final element in Mill's account. Mill believed that there were utilitarian grounds for holding that at least some autonomy rights should be *inalienable*—that is, that at least some autonomy rights generate limits on the rights bearer's autonomy to trade or surrender those very rights. For example, Mill argued that people should not be free to enter slavery contracts ([1859], 115).

Because Mill was a utilitarian, he qualifies as a *consequentialist* in the sense in which I use the term: A *consequentialist* about a given normative domain is someone who believes there is an explanation of that domain in terms of some measure (perhaps a distribution-sensitive measure) of nonmoral good. For Mill, the measure of nonmoral good was utility maximization. If the relevant measure is a (perhaps distribution-sensitive) measure of *well-being*, the view is *welfare consequentialist*, or *welfarist*. As a utilitarian, Mill was a welfare consequentialist about all of morality.

Utilitarianism is a maximizing view, which makes it a *teleological* view. My version of welfare consequentialism about human rights is not a maximizing view, because, on my account, both the amount and the distribution of well-being matter.[5]

Mill was the first person to attempt to give a consequentialist explanation of why governments should guarantee to all normal adults a package of robust, inalienable autonomy rights. I refer to this project as the *consequentialist project for autonomy rights*.

It is generally agreed today that Mill's attempt to carry out the consequentialist project for autonomy rights failed. One of the philosophers most responsible for this verdict is John Rawls. A little over 100 years after the publication of *On Liberty*, John Rawls wrote *A Theory of Justice*. Let me refer

to the author of that book as the early *metaphysical* Rawls, to distinguish him from the later *political* Rawls who would disavow parts of it.[6]

Because the early metaphysical Rawls was writing in the shadow of J. S. Mill, he began his book in a way that was best calculated to separate himself from Mill:

> Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust. Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. (1971, 3–4)[7]

The rhetorical force of his introduction tended to obscure how much the early metaphysical Rawls had in common with Mill—especially that metaphysical Rawls was attempting to bring to a successful completion the consequentialist project for autonomy rights.[8]

Unlike Mill, metaphysical Rawls did not try to give a consequentialist account of all of morality. His more modest goal was a consequentialist account of the justice or injustice of the basic institutions of society, especially the rights and duties established by law (including the constitution) and enforced by the coercive power of the state. Because our legal rights and duties provide a framework that defines our entitlement to the distribution of the benefits and burdens of social cooperation, metaphysical Rawls referred to his theory as a theory of *distributive justice*. Metaphysical Rawls believed that it was possible to develop the theory of distributive justice without working out a theory of *corrective justice* (i.e., a theory of the justice or injustice of punishment and other legal sanctions) if he focused on an ideal conception of justice for a society on the assumption of strict compliance with the just laws (1971, 8). On the assumption of strict compliance, it was not necessary for the early Rawls even to address issues of corrective justice.

The early Rawls's theory of distributive justice is almost universally identified with his two principles of justice, plus the priority rule that gives the first principle that specifies the basic autonomy rights (the Liberty Principle) lexical priority over the second (the Difference Principle) (1971, 302–303).[9] The Liberty Principle requires government protection of a package of robust, inalienable autonomy rights, very similar to Mill's package of rights, plus democratic rights. I refer to the combination of autonomy rights and democratic rights as *liberal rights*.[10]

Because Rawls's special theory of justice accords lexical priority to the rights covered by the Liberty Principle over considerations of well-being

(which are included in the Difference Principle), it is almost universally regarded as a nonconsequentialist theory of distributive justice, and the early Rawls himself seems to have regarded it as such (1971, 11). It is undeniable that the later *political* Rawls's (1993) account of the two principles (and the priority of the first over the second) is nonconsequentialist, but it is often overlooked that the early metaphysical Rawls regarded those two principles as a special case of a general conception of distributive justice that contained only a single, consequentialist principle, which I refer to as Rawls's *maximin expectation principle*—roughly to maximize the expectations of the least advantaged group.[11] Thus, for all his attempts to distance himself from Mill, if the early metaphysical Rawls had been successful, he would have succeeded in carrying out the consequentialist project for liberal rights, including both autonomy rights and democratic rights—that is, the project of providing a consequentialist explanation (in terms of his maximin expectation principle) of why governments should guarantee a package of robust, inalienable autonomy and democratic rights to all normal adults (i.e., a derivation of the Liberty Principle and of its lexical priority over the Difference Principle).[12]

Unlike Mill, the early Rawls never envisioned providing a consequentialist account of all of morality. He thought that the justification of liberal rights depended only on the theory of distributive justice. If I am correct that his theory of distributive justice was consequentialist, then, had he been successful, he would have successfully completed the consequentialist project for liberal rights.

Unfortunately, the early metaphysical Rawls's theory was flawed. There were two problems with Rawls's theory: (1) the inadequacy of his consequentialist maximin expectation principle as a principle of distributive justice and (2) the failure of his attempt to derive the lexical priority of the Liberty Principle (specifying the relevant autonomy rights) from his consequentialist maximin expectation principle. I discuss the first problem in chapter 4 and the second in chapter 7.

After the failure of Mill to successfully complete the consequentialist project for autonomy rights and the failure of the early metaphysical Rawls to successfully complete the consequentialist project for liberal rights, these consequentialist projects were largely abandoned. The later political Rawls himself disavowed his earlier attempt to give a consequentialist explanation of the lexical priority of the Liberty Principle over the Difference Principle and instead adopted the more promising line of giving it a nonconsequentialist explanation. Over the next 30 years, most of the influential accounts of rights and justice—those of the later Rawls (1993) as well as Barry (1995), Buchanan (2004a), G. A. Cohen (2008), R. Dworkin (2000), Habermas (1990 and 1996), Mills (1997), Nagel (1991), Nozick (1974), Nussbaum (2000), Sen (2009), and Thomson (1990)—were nonconsequentialist.[13]

Are Mill's and the early Rawls's consequentialist projects hopeless? There are many reasons for thinking that they are, especially if the account is

welfarist. The first reason is that such projects seem misguided. Consider, for example, autonomy rights. It seems almost perverse to try to ground a package of autonomy rights on considerations of (appropriately distributed) well-being, when there is a much more direct grounding of autonomy rights in the value of autonomy or consent. Most rights theorists today are some sort of nonconsequentialist, because most of them ground autonomy rights in the value of autonomy or consent, not well-being.

Even if the consequentialist were somehow able to give a consequentialist grounding to the same package of autonomy rights as the nonconsequentialist, the consequentialist explanation would be indirect and complex, whereas the nonconsequentialist account is simple and direct. This seems particularly true for a right against legal paternalism, which was an important element in Mill's package of autonomy rights. On a consequentialist account it would seem that there would be a strong presumption *against* any such right, because, after all, the goal of paternalistic laws is the promotion of well-being. But it seems obvious that a nonconsequentialist account based on the value of autonomy or consent would directly support a right against paternalism.

Thus, anyone who would seek to revive the consequentialist project for autonomy rights or for liberal rights takes on a substantial burden. It is not enough to rig together a Rube Goldberg consequentialist account that just happens to yield the same rights as a nonconsequentialist autonomy-based account. Because the consequentialist account is more complex than the nonconsequentialist account, it must do a better job than the nonconsequentialist account of explaining the contours of the relevant rights—for example, of the contours of an acceptable right against paternalism—than the nonconsequentialist account. The more complex consequentialist account can be favored over the simpler nonconsequentialist account only if the nonconsequentialist account generates explanatory problems that the consequentialist account is able to solve. Of course, the mere existence of explanatory puzzles does not discredit nonconsequentialist accounts, because all accounts have explanatory puzzles. However, it seems to me that there are a number of deep explanatory puzzles for nonconsequentialism that point to a deeper level of explanation, at which level the relevant explanatory principles are consequentialist.

Social Practice Consequentialism as an Explanatory Meta-Theory

In an earlier work (Talbott 2005) I explained my reasons for thinking that moral reasoning is largely bottom-up rather than top-down. Moral reasoning does not begin with principles that are self-evident or rationally intuited. Instead, our moral norms or principles are generally the product of millennia of experience and thought about actual or hypothetical particular cases.

Bottom-up moral reasoning is of two kinds: First, judgments about particular cases can provide support for principles or norms that explain them; second, judgments about particular cases can undermine principles or norms that are incompatible with them.

To explain my consequentialist account of rights, I need to distinguish two levels of moral thought. By *ground-level moral thought*, I mean the moral judgments and moral reasoning involved in a social group's shared practice of moral evaluation. It includes particular moral judgments of rightness and wrongness or justice and injustice ("The system of slavery practiced in the antebellum southern United States was wrong") and norms and principles ("Slavery is wrong"). Though I believe that the discovery of ground-level moral norms and principles is primarily a product of bottom-up reasoning, in a particular situation, ground-level moral reasoning can be either top-down, as when, for example, I conclude that slavery in the United States was wrong because I accept the principle that all human beings have a right not to be enslaved, or bottom-up, as when, for example, I conclude that human beings have a right not to be enslaved on the basis of studying the various institutions of slavery and deciding that each of them is wrong.[14]

Not all moral thought takes place at the ground level. There is another kind of moral thinking that philosophers sometimes do when they theorize about ground-level morality. This is explanatory reasoning, in which the goal is to explain ground-level moral thought. Not all explanations of ground-level moral thought qualify as *moral* explanations, because some explanations of them are debunking explanations. *Debunking explanations* explain ground-level moral thought in a way that implies that it is all a mistake. Thus, for example, a Marxist explanation of ground-level moral thought as a tool to promote the interests of the ruling class or an evolutionary explanation of ground-level moral thought as a "collective illusion of the human race, fashioned and maintained by natural selection" (Ruse 1995, 235) would be a debunking explanation.

In contrast, if an explanation of ground-level moral phenomena is not a debunking explanation—that is, if it at least leaves it open that some of the ground-level moral judgments are true (or morally appropriate)—I refer to it as a *moral meta-theory* and to the principles that it employs as *meta-theoretical moral principles* or *meta-level principles*. Meta-level moral principles are explanatory principles that attempt to explain the moral appropriateness of ground-level moral thought in a way that does not debunk it.

It is important to distinguish between ground-level consequentialism (*direct* consequentialism) and meta-level consequentialism (*indirect* consequentialism). Direct consequentialism has been pretty thoroughly discredited. There are many generally accepted particular moral judgments that conflict with almost any direct consequentialist principles (Nozick 1974, 28). Scanlon and Darwall have reinforced this objection to direct consequentialism by arguing that the concept of well-being itself plays almost no role in first-person moral reasoning (Scanlon 1998, 126–133) and by arguing that desirability (or

good consequences) is a reason *of the wrong kind* to warrant our second-person practice of making moral claims on others or of holding them accountable (Darwall 2006, 15, 104, 192, 311).

Both Scanlon and Darwall seem to take their arguments as arguments against moral consequentialism, but I think this is a mistake. The reason is that they do not seriously consider the kind of *indirect consequentialism* that uses consequentialist meta-principles to *explain* the moral appropriateness of ground-level moral thought (when it is morally appropriate), whether first- or second-person. To refute indirect, meta-level consequentialism, it is not enough to show that ground-level moral reasoning is not consequentialist. It is necessary to consider whether there is a consequentialist meta-theory that explains the moral appropriateness of the nonconsequentialist ground-level moral reasoning.[15]

For example, Brandt (1992) believed that ground-level moral reasoning should be guided by simple *nonconsequentialist* moral rules (e.g., that lying is wrong), but he thought that the meta-principle that *explained* why ground-level moral reasoning should be guided by such rules was a consequentialist (rule utilitarian) one that favored systems of rules that maximized utility. Brandt's account is an indirect consequentialist account, because it uses consequentialist meta-principles to explain the moral appropriateness of nonconsequentialist first-order moral principles.

Similarly, Mill's [1859] theory of robust, inalienable liberty rights was also a consequentialist meta-theory, as any plausible consequentialist account of *robust* rights would seem to have to be. Mill's brand of indirect consequentialism was more general than Brandt's, because it was not limited to explaining the justification of systems of rules, but could be applied to explain the justification of any social institution or practice (e.g., the family), whether or not it could be defined by a system of rules. There is no rule book for being a good parent, nor is it plausible that there could be one. But Mill's consequentialist meta-theory could easily be used to explain the justification of the family as a social practice.[16] I refer to this kind of indirect consequentialism as *social practice consequentialism.*

Mill proposed a consequentialist meta-theory for all ground-level moral thought. As I interpret him, the early metaphysical Rawls (1955 and 1971) had a consequentialist meta-theory, but the meta-theory addressed only distributive justice, not all of morality.

It is an interesting question whether the early Rawls himself thought of the maximin expectation principle as a moral meta-principle or as itself a part of ground-level thought about justice. However, there is no doubt that Rawls thought that the constitutional constraints on legislators and judges would not be consequentialist, but would be given by principles establishing the lexical priority of the rights covered by the first principle of justice, autonomy rights and democratic rights. In a legal system in which judges applied the maximin expectation principle in their decisions, judges would make exceptions to laws whenever they thought it would maximize the expectation

of the least advantaged group to do so. This is not Rawls's view of the role of judges (1971, 196–201).

If Mill proposed a consequentialist meta-theory of all of morality and the early Rawls's theory can be taken to be a consequentialist meta-theory of distributive justice, then my consequentialist meta-theory is somewhere in between the two. In this volume, I attempt to articulate a consequentialist meta-theoretical principle that explains not all of ground-level morality, but only a part of it. But that part turns out to include all of what Rawls thought of as distributive justice. So my consequentialist meta-theory is narrower than Mill's, but broader than the early Rawls's. However, the list of rights that I defend is more expansive than Mill's list of autonomy rights and more expansive than Rawls's list of liberal rights. I refer to them as *human rights* because they are the robust, inalienable rights that all governments should guarantee to all their citizens. Because my consequentialist principle explains the content of those human rights norms, I think of it as providing a consequentialist explanation of human rights.

Primary and Secondary Ground-Level Moral Thought

To classify my kind of meta-level consequentialism, it is necessary to say something more about ground-level morality. For my purposes, it is useful to divide ground-level moral thought into two parts: primary moral judgments and secondary moral judgments. Examples of *primary* moral judgments, norms, and principles are ordinary judgments about the rightness or wrongness of particular actions or kinds of actions—for example, that murder is wrong. The *secondary* moral judgments are moral judgments about the enforceability of other moral judgments—for example, judgments of the permissibility of self- and other-defense, deterrence, and punishment.[17] The judgment that murderers may be imprisoned is a secondary moral judgment. There is an infinite hierarchy of secondary moral judgments. At the first level are judgments about the enforceability of primary moral judgments (e.g., that murderers may be imprisoned). At the next level are judgments about the enforceability of secondary moral judgments on the enforceability of primary moral judgments (e.g., the judgment that a convicted murderer may be punished for attempting to escape from prison). There is no theoretical limit to the number of levels of secondary moral judgments (e.g., the judgment that a person imprisoned for attempting to escape from prison should be further punished for further attempts to escape), but in practice, the number is quite limited.

Because Mill attempted a meta-level consequentialist explanation of all of morality, he assumed the burden of providing a consequentialist explanation of both primary and secondary ground-level moral thought. I am sympathetic to this project, but it is much too large a project for me to take on here. In this book, I limit my explanatory project to primary ground-level

moral judgments. And even here, my project is limited. Although I am sympathetic to the project of providing a meta-level consequentialist explanation of all of ground-level primary moral judgment, even that project is too large for me to take on here. My more modest project is to explain the moral appropriateness of certain improvements to primary ground-level moral thought.

Let me explain. If they persist long enough, all moral traditions change over time. I believe that, at some point in the history of any moral tradition, the moral appropriateness of at least some changes in its primary moral judgments (when they are appropriate) is explained by a consequentialist meta-principle. When a moral tradition has passed this point in its history, I will say that it has crossed the *consequentialist threshold*. Once a moral tradition has crossed the consequentialist threshold, the moral appropriateness of most changes in its ground-level primary moral judgments is explained by a consequentialist meta-principle, which provides a sufficient condition for moral improvement. Because this consequentialist meta-principle turns out to be the main meta-principle in the explanation of the moral appropriateness of human rights norms, I refer to it as the *main consequentialist meta-principle*, or the *main principle* for short.[18]

I have no way of determining exactly when a moral tradition crosses this consequentialist threshold, but every major religious and moral tradition has crossed it. One positive test for whether or not a tradition has crossed this threshold is whether it endorses some version of the Golden Rule.[19] Every major religious and moral tradition has done so.[20]

Why does the main principle come into play only after a moral tradition has crossed the consequentialist threshold? The guiding idea is this: Initially, moral practices are favored in processes of biological and cultural selection for their advantages. We now know that what seem to be moral or proto-moral practices have even been selected for in nonhuman species (e.g., de Waal 2006). At these early levels of moral development, it may be that evolutionary constraints set the standards for improvement in ground-level moral practices. But when a culture reaches a certain level of moral development, those who receive moral training in it acquire a kind of moral sensitivity that replaces imperatives of biological and cultural selection in influencing changes in ground-level primary moral practices. I say something more about this moral sensitivity in chapter 5. There is no way to tell precisely when this transformation occurs, but when a cultural tradition adopts a version of the Golden Rule, we can know that it has occurred. Although the Golden Rule itself is not a rule of reciprocity (as it would be if it enjoined us to love our friends and hate our enemies), when it is adopted as part of a shared moral practice within a social group, it functions to establish mutually beneficial reciprocity relations. Individuals do not usually benefit directly from acting on the Golden Rule, but everyone benefits from other people's willingness to act on it. The main principle is a principle for making moral improvements in a system of moral reciprocity relations.

What about secondary moral practices—those that have to do with enforcement (e.g., norms of punishment)? It is sometimes claimed that there is a version of the Golden Rule that justifies retributive punishment of those who violate primary moral norms (e.g., an eye for an eye, a tooth for a tooth). Call this the *Retributivist Golden Rule.* The Retributivist Golden Rule seems clearly mistaken. Even setting aside its morally problematic implications (e.g., that the proper punishment for rapists is to be raped), it is clearly inadequate as a secondary moral principle. Consider the crime of stealing $1,000. It is easy to see that the appropriate punishment for that crime may be much more than restitution and a fine of $1,000, because such a fine would not effectively deter stealing if the probability of getting caught was less than one-half. Thieves would make money if they just regarded having to pay the fine when they got caught as one of the costs of doing business.

Even though the Golden Rule does not seem to apply to secondary moral thought about punishment of the guilty, it seems to me quite plausible that secondary moral thought can be explained by a consequentialist meta-principle. However, to try to carry out the explanation is beyond the scope of this book. For present purposes, it is necessary to narrow the focus to the project of providing a meta-level consequentialist explanation of improvements in ground-level primary moral thought, for moral traditions that have crossed the consequentialist threshold.

Rawls was able to make a clean division between primary and secondary moral thought by assuming strict compliance with his two principles of justice (1971, 8). If there is strict compliance, enforcement is unnecessary, and so a meta-level theory of the enforceability of moral judgments is also unnecessary. However, I do not adopt Rawls's extremely idealized assumption of strict compliance. I discuss the justification of moral norms and principles, and, especially, human rights, in more realistic cases in which it is known that enforcement will be necessary. In such cases, it is often thought that consequentialists must allow for legal systems that knowingly punish the innocent (Nozick 1974). In chapter 6, I argue that everyone, consequentialist or nonconsequentialist, must allow for legal systems that are known to punish some innocent defendants (because it is inevitable that some innocent defendants will be mistakenly judged to be guilty), though neither is committed to endorsing a system that punishes defendants known to be innocent.

My main focus will be on ground-level moral thought about human rights and, by extension, the constitutional provisions or laws that guarantee them. Included in ground-level moral thought are the rationales that judges give to justify their legal decisions, when those decisions overrule prior law or apply old law to a new kind of case. Also, included are the rationales that legislators give to justify constitutional amendments or to justify laws, when the considerations are considerations of justice, rather than considerations of how best to promote the interests of their constituents. When judges' or legislators' rationales involve considerations of justice or fairness, they are a part of ground-level moral practices that the main principle applies to.[21] It is

important to keep in mind that when I say that the main principle applies to changes in ground-level moral thought, I mean to include this kind of legal thought, also, because, in this way, the main principle is the most important principle for explaining the appropriateness of changes in human rights. The main principle does not apply to secondary norms, so, in all the examples I discuss, I just assume that the enforcement provisions of the relevant laws satisfy the relevant proportionality constraints.

Given these preliminaries, I can simplify my exposition by assuming, unless I say otherwise, that by *ground-level moral and legal thought* I mean changes in ground-level primary moral and legal thought in a moral tradition that has passed the consequentialist threshold. That is the ground-level moral and legal thought that my consequentialist meta-theory aims to explain the moral appropriateness of.

My Explanatory Strategy

Consequentialists typically begin by defining the important terms (e.g., *well-being*) and then offering some formula for rightness or justice in terms of well-being (e.g., in terms of maximizing overall well-being). I don't have direct rational insight into self-evident truths about morality and justice, so I can't define any of the most important terms that I use and I cannot provide a precise formula for rightness or justice.

Particular Moral Judgments

My approach is to work primarily in the other direction, bottom-up rather than top-down. As I see it, ground-level moral principles (including principles of human rights) are the result of a largely bottom-up process of discovery, based on ground-level *particular moral judgments*—that is, moral judgments about actual and hypothetical particular cases (e.g., that Hitler's attempt to exterminate the Jews was wrong). It is important to understand how we can discuss particular cases. Here is an example: Typically, in discussing particular cases, I assume that an increase in life expectancy represents an increase in well-being. In such cases, I am depending on your ability to imagine cases in which it does increase well-being, because I know that, for any interesting generalizations that I might formulate, there will almost always be exceptions. It is not always true that increases in life expectancy increase well-being. For example, we can easily imagine cases in which someone with a fatal disease faces a short period of suffering that will end with death. Prolonging their period of suffering would not generally be a way of promoting well-being. Notice that, even here, I am relying on your ability to imagine cases of the relevant kind, because I am not denying that there are exceptions to the exception—that is, cases in which it would promote well-being to extend the period of suffering—for example, if living for 2 days

more would allow time for to reconcile with an estranged family member, something that was very important to the suffering person, before dying.

How are we able to refer to the relevant kinds of examples with finite descriptions, if adding further information can change our moral judgment about a particular case? This is a fascinating question that deserves more attention than I can give it here. Part of the answer is that we do it by making our intent clear to our audience (e.g., the intent to describe an example of an increase in well-being) and then providing the audience with enough information for them to be able to imagine the relevant kind of example (e.g., making available a drug that significantly increases life expectancy). When my audience knows that I intend to describe an example involving an increase of well-being and then I ask the audience to imagine a case in which a certain drug increases life expectancy, the audience will look for examples of drugs that increase life expectancy in a way that increases well-being. If such examples are difficult for the audience to find, then I did not provide enough information. But if the information that I provided makes such examples easy for the audience to find, then there is no need to provide more information. Providing more information might rule out some exceptions, but there is no need to think that we must be able to describe examples in a way that rules out all exceptions, in order to be able to discuss particular cases.[22]

Ground-Level Moral Norms and Principles

How can we explain particular moral judgments? The simplest kind of explanation would be an explanation in terms of ground-level moral generalizations— that is, ground-level norms or principles. For example, the norm "coercion is wrong" would explain the variety of particular cases involving wrongful coercion. But what about cases, hypothetical as well as actual, in which coercion is not wrong? J. S. Mill gives the example of a person about to cross an unsafe bridge ([1859], 109). If there is not time to explain the danger, the use of force is permitted to stop that person from crossing the bridge.

Typically, the first reaction to the discovery of exceptions to ground-level moral norms is to try to fix the norm by building exceptions into it or by finding a more general ground-level principle that explains why the norm holds in those cases in which it does and why it fails to hold in the exceptional cases. For example, one might propose a new ground-level norm: Coercion is wrong unless necessary to prevent death or serious injury in a case in which there is no time to explain why there is a danger of death or serious injury. Or one might instead seek a more general ground-level principle, and hit upon a version of the Golden Rule: Do unto others as you would have them do unto you. This version of the Golden Rule would prohibit most cases of coercion, but would allow an exception in the case of the unsafe bridge.[23]

One of the great puzzles of moral philosophy is that this process of adding exceptions to our ground-level norms or finding new ground-level norms to cover the exceptional cases so far discovered does not ever seem to end. There

are always more exceptions to the ground-level norms or principles. This result is so familiar that Scanlon simply assumes that the moral principles he discusses are actually "labels for much more complex ideas" (1998, 199) that cannot be captured in a simple rule. Because of the potential for exceptions to a given principle, exceptions to the exceptions, and so forth, Scanlon thinks there must be an indefinite number of moral principles (201). Dworkin makes the same point about principles in the law: They all have exceptions (1977, 25).

But if ground-level moral and legal norms and principles typically have exceptions, there is no complete explanation of a particular moral judgment at the ground level. The reason is simple. If by "coercion is wrong" we understand "coercion is usually wrong," then the norm cannot by itself explain the wrongness of a particular act of coercion, because the full explanation would require not only the norm that coercion is usually wrong but also an explanation of why the relevant particular case is one of the "usual" rather than the "unusual" cases.

Meta-Level Moral Principles

Of course, it may be that there are true exceptionless ground-level norms or principles that explain all the true ground-level particular moral judgments and our problem is just that we have not yet discovered them. The alternative that I want to seriously consider is that there is a higher level of explanation at which it is possible to explain the moral appropriateness of ground-level moral judgments, including particular moral judgments, norms, and principles (when they are appropriate) and their moral inappropriateness (when they are inappropriate).[24] Surprisingly, at the meta-level, we discover an exceptionless principle that not only explains the moral appropriateness of changes to the ground-level particular judgments, norms, and principles, but, as I show in chapter 5, it also explains why substantive ground-level norms and principles always (or almost always) have exceptions. I call this meta-level consequentialist moral principle the *main principle*. The main principle explains the moral appropriateness of most *changes* in ground-level primary particular moral judgments and ground-level primary norms and principles (when they are morally appropriate) in moral traditions that have passed the consequentialist threshold.

Puzzles about What Is to Be Explained

I have said that the main principle explains the moral appropriateness of most changes in ground-level moral and legal thought (when they are morally appropriate). For ease of exposition, let's focus on moral thought. The application to legal thought is exactly parallel. When we ask how we can test such a theory, a deep puzzle emerges. One way to test the theory would be to look back on the history of ground-level moral thought to identify the cases

in which exceptions have been made to accepted ground-level norms or principles. But this would be a fallacious test, because a moral meta-theory is not a descriptive theory. It is not an attempt to explain all of the changes in ground-level moral thought that have actually occurred. It is an attempt to explain the moral truth or appropriateness of those changes that were morally true or appropriate (and moral falsity or inappropriateness of those that were morally false or inappropriate).

So it seems that we must test the theory against our own considered judgments about which of the changes were morally appropriate—or, to be more exact, which were moral improvements—and which were not. This is a cause for worry. What is to keep me from adjusting my judgments about which moral changes have been improvements to fit my theory?

In addition, it would seem that any such a theory would be hopelessly relativistic, because there is so much disagreement about which moral changes have constituted improvements. I regard the extension of equal rights to women as an important moral advance, but the Taliban regards it as an example of moral degradation. Who is to decide which changes qualify as improvements? In this book, I do not maintain neutrality between different views of moral progress.[25] But the test of my theory is not that it persuades me or people who share my beliefs about moral progress. As I explain in chapter 7, the best test of both my theory and the Taliban's theory (though not an infallible one) is how they fare in the process of free give-and-take of opinion. Of course, if the Taliban had their way, they would suppress the process of the free give-and-take of opinion. They could thus prevent any challenges to their claim that their theory was justified. But claiming it would not make it so. I discuss these issues more fully in chapters 7 and 14.

There is another problem, also. Any adequate theory of moral improvement must have implications that go beyond the actual changes that have occurred in the past. It will have implications for which potential future changes would be moral improvements and which would not. Are we supposed to test these implications against our *current* judgments of which future changes would be moral improvements and which would not?

That cannot be a satisfactory test. When we look at the past history of changes that we now regard as moral improvements, we find that there were times when most people had a moral blind spot that prevented them from recognizing that the change would be a moral improvement. For example, very few Europeans raised moral objections to the slave trade in the sixteenth century. Even in the eighteenth century, the slave trade flourished and some of the authors of the Declaration of Independence were able to hold that all men are created equal while also defending slavery. Given the prevalence of moral blind spots in the past, it would be a display of *hubris* to think that we ourselves don't also have moral blind spots. But if we have moral blind spots, then there are some changes to our own ground-level moral thought that would be moral improvements, but, due to our own moral blind spots, we don't realize that they would be.

This has the following paradoxical implication: Suppose someone articulated a moral meta-theory that compellingly explained the moral appropriateness of past changes in ground-level moral thought now regarded as moral improvements and also identified exactly those changes in current ground-level moral thought that would now be regarded as moral improvements. Such a theory might be very useful for many purposes, but we would be almost certain that it was false, because it would fail to identify our own moral blind spots. A fully adequate moral meta-theory must identify some potential improvements in ground-level moral thought that we would not today regard as improvements.

Thus, there is no adequate synchronic test of a moral meta-theory. A moral meta-theory must be tested, in part, diachronically, by the way that ground-level moral thought changes in the future. Because future changes in ground-level moral thought can themselves be influenced by our moral meta-theories, a moral meta-theory theory can be tested not only by its predictions about changes that will in the future be regarded as moral improvements, but also by the changes in ground-level moral thought that it contributes to bringing about.

Another way of putting this point is to say that a moral meta-theory is a theory of past changes in ground-level moral thought that have been improvements and of potential future changes that would be improvements. The puzzle is that we must test such theories by our own ground-level moral thought, which we have good reason to believe is itself subject to improvement. Of course, if our ground-level moral thought is massively mistaken, our moral meta-theories will be massively mistaken also. However, I do not mean to be raising skeptical worries here.[26] A moral meta-theory that provided a satisfactory explanation of the moral appropriateness of past changes in ground-level moral thought that we now take to have been improvements and a satisfactory explanation of the moral appropriateness of potential future changes that future generations will come to regard as improvements would be a stunning accomplishment. That would not only be good evidence for its truth, it would be the best possible evidence for its truth.

Improvements Are Comparatively Better, Not Optimal

Because my goal is to explain moral improvements, my consequentialist theory is not an optimizing theory. It is almost certain that no human society will ever discover an optimal moral system, on any reasonable criterion of optimality (cf. Sen 2009). But all human societies can improve their moral practices. The principle that explains which changes are improvements has to make comparative evaluations of only a relatively small number of relevant alternatives—usually, the *status quo* and one or two potential changes to the *status quo.* It is much more likely that human societies could satisfy such a comparative principle than that they could ever satisfy any plausible optimizing principle.

Strict Universality of Particular Moral Judgments and of Meta-Level Principles

One of the most surprising claims in my first volume (Talbott 2005) was that by reasoning in a largely bottom-up manner it is possible to discover fundamental moral principles that are strictly universal—that is, true of all rational beings in all possible worlds. Traditionally, it was thought that the only way to have knowledge of strictly universal (i.e., metaphysically necessary) principles of any kind was through direct *a priori* insight. Because I don't claim to have any direct *a priori* insight, it is surprising that I would claim that there are strictly universal moral meta-principles and that we are engaged in an ongoing historical process of trying to figure out what they are.

The key to understanding how it would be possible to discover such principles is to understand that our true particular moral judgments are also strictly universal, though in a slightly different sense. When I make a particular moral judgment (e.g., that it was wrong of the Western European colonists to enslave American natives or to force them to adopt the Christian religion) I do not claim to be infallible. But I do think that we are justified in placing a great deal of confidence in them in clear cases. These judgments are largely true, and when they are true, they are objectively true. They are true not just for human beings or for those who share our moral tradition. When they are true, they are true for any rational being. This is the sense in which particular moral judgments can be strictly universal. If this is right, then we can use particular judgments about actual and hypothetical cases to support principles that apply to actual and hypothetical cases. Were we to discover the fundamental principles that explain all actual and hypothetical cases, they would be true in all possible worlds. So they would be strictly universal principles.

For most of human history, the goal of moral inquiry has been to formulate exceptionless ground-level moral principles. The failure to do so has led many philosophers to deny that there are any (Dancy 2004). It is somewhat surprising that there might be a meta-level explanation of why exceptionless ground-level moral principles are so rare and even more surprising that the meta-level explanation would employ an exceptionless meta-theoretical principle. But there is and it does, as I explain in chapter 5. So it turns out that there are strictly universal moral principles, but they are meta-theoretical principles, not ground-level principles.

Contingent Universality of Human Rights Norms or Principles

Because human rights norms or principles are ground-level norms or principles, we should not expect them to be exceptionless. I have already acknowledged that they are not, when I said that the project is to explain human rights that are *robust* but not *absolute*. Some readers will be disappointed by

this admission. They will not want to give up on the search for exceptionless ground-level human rights principles. I hope that the explanation of why it is almost inevitable that ground-level principles have exceptions in chapter 5 will help to reconcile those readers to this result.

If ground-level human rights principles are not exceptionless, then they are not strictly universal, not true in all possible worlds. This should not be surprising. Human rights depend on human capabilities. In possible worlds in which human beings had very different capabilities, they would be expected to have very different rights.

The universality of human rights is not strict universality, but it is an important kind of contingent universality: Given what we know about human beings and human societies in this world, the main principle explains why it is morally appropriate that *all* human societies guarantee autonomy rights (and other rights on my list of human rights) for *all* normal adults. This is the sense in which the rights on my list of human rights *should be universal.*

Justifying Government Coercion

Although the main principle applies to exceptions to any ground-level primary moral norm or principle, a particularly important category of exceptions is the category of exceptions to the ground-level moral prohibition on coercion, because that is the prohibition to which judges and legislators must be able to justify exceptions if they are to be able to justify making new law and modifying old law. Laws are coercive. Because coercion is generally wrong, the rationale for coercive laws must state an exception to the general rule against coercion. To a first approximation, the main principle supports exceptions to the general rule against coercion when such exceptions, evaluated as a social practice, equitably promote the well-being of those who are coerced. There are two important kinds of laws that can be used to promote well-being:

(1) Paternalistic laws. These are laws that limit a person's liberty for her own good, even though the person herself may disagree. In chapters 12 and 13 I argue that when certain basic rights are guaranteed, normal human adults should have liberty rights to freedom from government paternalism, unless it satisfies a special kind of hypothetical consent standard, the most reliable judgment standard. The most reliable judgment standard is a ground-level standard. The explanation of its moral appropriateness is a meta-theoretical explanation that employs the consequentialist main principle. So the standard for rights against paternalism that I articulate in chapter 12 is not consequentialist, but the explanation of why that standard is morally appropriate is.

(2) Legal solutions to collective action problems (CAPs).[27] This is the most important category of laws promoting well-being. CAPs are ubiquitous. Climate change, pollution, and fisheries destruction are negative examples,

in which the outcome is bad if everyone drives gas guzzlers, pollutes, or overfishes, but in which an individual person's contribution to the badness of the outcome is so small as to be negligible and each individual has a reason to do the slightly bad act, because it is more costly to her not to do it. Fire protection, highways, traffic signals, and medical research are all positive examples, in which the outcome is good if everyone contributes, but each individual's contribution to the good outcome is so small as to be negligible and each individual has a reason to avoid doing the slightly good act because it is more costly to her to do it. I refer to these problems as *N-Person Prisoners' Dilemmas*.[28] In such situations, by convention, the act that leads to the better results when chosen by everyone is called *cooperating*. The act that leads to the worse results when chosen by everyone is called *defecting*. A quick test for an N-Person PD is whether there would be some temptation to free ride— that is, to defect if everyone else or almost everyone else is cooperating. This test shows that typical cases of stealing, cheating, lying, promise-breaking, even murder, also generate an N-Person PD. Those who steal benefit from others' not stealing to be able to enjoy the benefits of their theft.

CAPs are not only ubiquitous, but they are what might be called *productive*: A solution to one can and often does generate others. Market economies are a solution to a CAP (the productive investment CAP), but they themselves generate possibilities for corporate fraud and market bubbles, both CAPs. Governments are a solution to CAPs, but voting itself is a CAP. Not all CAPs should be solved. The main principle explains why. Price fixing is a solution to a producers' CAP that the main principle does not endorse.

Human societies could not thrive unless they had ways of solving CAPs. Indeed, human societies would probably not exist were it not for CAPs. It is the existence of CAPs that gives an evolutionary advantage to social species, such as human beings (Wright 2000). Recognizing and sanctioning cheaters and other free riders is so important to a social group that evolution has almost surely endowed us with the psychological equipment to detect cheaters and respond appropriately to them (Cosmides and Tooby, 1992). Morality itself is one social practice that helps to solve CAPs. A legal system is another.

Hobbes [1651] thought that life without a government to make and enforce laws would be so awful that any government, no matter how bad, would be infinitely better than no government at all. That is an exaggeration, but it is not an exaggeration to say that solving CAPs is the most important function of a government. Governments implement coercive solutions to CAPs by punishing defectors. They imprison murderers and thieves, fine polluters, establish fire departments, fund medical research, establish and administer a system of police, courts, and prisons to enforce their laws, and punish those who don't pay their taxes to pay for all these solutions to CAPs.

In this book I argue that, when established against a background of the other basic rights, part of the rationale for constitutionally limited democratic rights is their role in solving CAPs in a way that equitably promotes

well-being. The ground-level principles establishing democratic rights are nonconsequentialist (e.g., one person, one vote). It is at the meta-level, where the rationale for ground-level democratic principles is consequentialist—that establishing such rights is a good way of solving CAPs in a way that equitably promotes well-being (at least, better than the other relevant alternatives). As I explain in chapter 10, the main principle endorses democratic rights, in combination with constitutional protections for robust, inalienable human rights, because of their tendency to equitably promote well-being.

What Is Normative Truth?

I believe that the main principle is a *true* meta-level moral principle. To say this is to say that some changes in ground-level moral practices *really are* improvements. They are not just improvements from a liberal point of view or a cultural or religious point of view or a species point of view. They are improvements from the objective point of view (Nagel 1986).

To believe in normative truth, it is not necessary to believe in any weird entities or forces. All that is required is to believe that there can be *real* moral progress. If there are true moral meta-principles, they are not written on stone tablets. What kind of truths are they? In a sense, this entire book is an extended answer to that question.

Moral truths are only one category of normative truth. There are also normative truths about what it is rational to believe and truths about what it is rational to do in nonmoral situations. In all these cases, it is very difficult to articulate exceptionless principles, but not so difficult to describe some clear examples of rational and irrational belief or rational and irrational action (in nonmoral contexts) or moral or immoral action. In all three of these cases, if there are truths, they are not purely descriptive truths about what people actually do or believe, but truths about what any rational or moral being should or should not do or believe. Anyone who believes in normative truth is a *normative realist.*

Some people think that there could be normative truths only if God made them true (e.g., E. O. Wilson 1998, chap. 11). These people must think there are no objective normative constraints on what God believes or does. This is a puzzling view. Could God have made it true that it was rational to believe all contradictions—for example, that God exists and does not exist or that God is omnipotent and not omnipotent? Could God have made it true that it is morally right to torture innocent children merely for fun? These questions are puzzling enough to motivate our taking seriously the possibility that normative truths are constraints on all rational beings, including God.

Some people are moral noncognitivists. They think there aren't any normative truths (e.g., Gibbard 1990). Noncognitivists think that our normative avowals evince a certain kind of attitude. There is nothing objective for those avowals to correspond to or to fail to correspond to. These views are close

relatives of the views of those who think that God makes normative truths. If there is no God to make them *true (or false)*, then it is human attitudes and emotions that make them *appropriate (or inappropriate)*. One way to be a normative realist is to believe that human attitudes could not make the law of noncontradiction appropriate or inappropriate and could not make torturing children merely for fun appropriate or inappropriate.[29]

But really, I can hear someone say, isn't it enough to find normative principles that apply to all human beings? Why think there are any strictly universal normative truths that apply to all rational beings? To answer these questions, it is useful to consider the example of utilitarianism.

What Universal Moral Truths Might Be Like

Though the utilitarian principle of maximizing overall (i.e., total or average) well-being is not an adequate principle of morality, either as a ground-level or a meta-level principle, it is close enough to give us some idea of how the bottom-up process of moral inquiry might lead us to a strictly universal moral principle. It really does seem that well-being is something that would be important to any rational being. The problem with utilitarianism was that it mistakenly concluded that maximizing overall well-being would be important to any rational being. As Rawls (1971) pointed out, institutions that maximize well-being need not promote everyone's well-being. Thus, utilitarianism allows for the possibility of reducing some people's well-being in order to produce a more-than-offsetting increase other people's well-being. This is utilitarianism's distributional blind spot. The main principle does not have a distributional blind spot. It aims at equitably promoting everyone's well-being. Distribution matters.[30]

Why does the main principle apply to all rational beings? Consider only one application of the main principle: determining morally appropriate norms for solving CAPs. CAPs are practically unavoidable for rational beings who interact with other rational beings.[31] And even rational beings who never find themselves in a collective action problem with other rational beings could still ask themselves what they should do if they ever were to find themselves in such a situation. So we should at least entertain the possibility that there might be principles that determine the moral appropriateness of norms for solving CAPs for any kind of being. If so, they would be strictly universal moral meta-principles. The possibility of such principles should not be ruled out at the beginning of our inquiry.

The Main Principle and Human Rights

In the previous volume, I outlined nine basic human rights. I called them *human* rights, because they are rights that should be guaranteed to all normal

human beings and *basic* because they must be guaranteed for a government to meet a minimum standard of moral legitimacy.[32] Here is the list of the basic human rights:

1. A right to physical security
2. A right to physical subsistence (understood as a right to an opportunity to earn a subsistence for those who are able to do so and a welfare right for those who are not)
3. Children's rights to what is necessary for normal physical, cognitive, emotional, and behavioral development, including the development of empathic understanding
4. A right to an education, including a moral education aimed at further development and use of empathic understanding
5. A right to freedom of the press
6. A right to freedom of thought and expression
7. A right to freedom of association
8. A right to a sphere of personal autonomy free from paternalistic interference
9. Political rights, including democratic rights and an independent judiciary to enforce the entire package of rights

It is useful to group these rights into a small number of partially overlapping categories. I refer to the first eight items on the list as *autonomy rights*, because they are the rights that are necessary for citizens to develop and exercise their autonomy.[33]

The eight autonomy rights can be further divided into *development-of-judgment* rights (the first four rights on the list), because they are necessary to develop the capacity for good judgment (the ability to make reliable judgments about one's own good) and *exercise-of-judgment rights* (the next four rights on the list), because they are necessary for someone who has the capacity for good judgment to actually have good judgment and exercise it. I have more to say about some of these rights in coming chapters.

The final item on the list, political rights, is necessary to make governments appropriately responsive to the judgments of their citizens. An independent judiciary is necessary to protect all of the items on the list from government abuse or majority tyranny.

In the previous volume, I discussed both consequentialist and nonconsequentialist rationales for the nine basic rights, without choosing between them. In this volume I choose. I believe that the consequentialist main principle is the best meta-level explanation of why governments should guarantee their citizens the nine rights on my list. For a government to reliably promote the (appropriately distributed) well-being of its citizens, it must guarantee the nine basic rights on my list.

In this volume, I discuss some of the basic human rights in more depth—security rights (chapter 6); a right to freedom of thought and expression and the

related right of freedom of the press (chapters 7 and 8); democratic rights (chapter 10); and a liberty right against legal paternalism (chapters 12–13). In addition, I identify five further kinds of human rights—that is, robust, inalienable rights that should be universal:

10. Economic rights (chapter 9)
11. Negative opportunity rights—that is, rights against discrimination (chapter 11)
12. Positive opportunity rights—rights to certain capabilities (chapter 11)
13. Social insurance rights (chapter 11)
14. Privacy rights (chapter 13)

Is my list of human rights too long? It would be too long if my goal were to identify the rights on which there currently exists an international overlapping consensus. However, whatever international consensus exists today leaves lots of room for improvement. My list is intended to point to what the consensus *should be*, and to explain why.

How to Make a Case for Consequentialism at the Meta-Level

In the competition between consequentialist and nonconsequentialist theories of human rights, the consequentialist starts out at a distinct disadvantage. The ground-level human rights principles themselves are nonconsequentialist, so they seem to invite a nonconsequentialist explanation. In addition, one of the most important categories of human rights on my list is the category of autonomy rights, and it seems almost self-evident that a nonconsequentialist explanation of autonomy rights in terms of the importance of autonomy would be a simpler and more direct explanation than a consequentialist explanation in terms of equitably promoting well-being.

It turns out that the simplicity of the nonconsequentialist account is also its Achilles heel. To compare the two accounts, we must consider not only how directly and simply they explain the relevant categories of rights, but also how well they are able to explain the contours of the rights in each of the categories. Because the contours of the individual rights involve many nuances and irregularities, a simple theory has difficulty in adequately explaining them. I will try to show that a consequentialist account does a much better job.

At the most fundamental level of analysis in a theory, one finds the central concepts of the theory. In nonconsequentialist theory, two of these fundamental concepts are autonomy and consent. In this book, I try to show that there is an even more fundamental, consequentialist, meta-theoretical level of explanation at which level we can explain the moral significance of these concepts. I discuss the significance of consent in chapter 9 and the nature and significance of autonomy in chapters 12 and 13.

One way of trying to cast doubt on nonconsequentialist theories of human rights is to raise puzzles for them, especially puzzles that seem to have a consequentialist solution. So I raise lots of puzzles in this book. For example, in chapter 6, I show how, in theory, it could be a moral improvement to do away with punishment altogether and how, in theory, a move to a system of strict criminal liability could be a moral improvement. In chapter 8, I show that Rawls's and Habermas's theories fail to support a constitutional right to freedom of expression that includes the expression of intolerant subversive advocacy. In chapter 10, I show how, in theory, we could be warranted in replacing democratic elections with a system of deliberative polling, and I provide a consequentialist solution to the puzzle of why any rights should be inalienable. In chapter 11, I raise a puzzle for views, such as Dworkin's (2000), that are based on the distinction between brute luck and option luck.

In the book, I give extended critical consideration to many of the most influential nonconsequentialist theories, including those of Nozick (in chapters 2–3), Thomson (in chapter 4), Rawls (in chapters 7 and 10), Habermas (in chapters 7 and 10), Dworkin (in chapter 11), and Feinberg (in chapter 12) and briefer critical consideration of many others.

It is important not to overstate the significance of the puzzles I raise for nonconsequentialists. As Kuhn (1962) observed about scientific theories, every theory has its puzzles. One way to allay doubts about my consequentialism is to address and resolve some of the well-known puzzles for consequentialism. So I do, including many of the standard objections to theories based on well-being in chapter 4, Nozick's Wilt Chamberlain example in chapter 3, examples of punishing the innocent and organ harvesting in chapter 6, the problem of seeming to justify lots of paternalism in chapters 12 and 13, and a host of objections in chapter 14. But I could never eliminate all puzzles for consequentialism. In philosophy, every theory has its puzzles.

For that reason, in deciding among theories, the decision often comes down to such considerations as the way a theory unifies disparate phenomena, illustrated by the way that my account in chapter 5 provides a unified explanation of the defeasibility of moral and legal reasoning (and, potentially, all reasoning); or the way that my account of the role of tort law in chapter 9 unifies a market economy and a system of tort law into a single self-regulating system; or by the way the main principle in chapter 3 and the Millian epistemology in chapter 7 unify the seemingly disparate rights on the U.N. Universal Declaration of Human Rights, as explained in chapters 6–13.

To many people, it seems obvious that the moral appropriateness of human rights norms could not be based on their contribution to equitably promoting well-being. Even worse, they see it as a threat to the human rights movement to even suggest that it might be. These people realize that the most influential arguments *against* human rights are typically based on well-being—for example, that poor countries can't afford human rights because they need to

encourage economic development. Because considerations of well-being are usually used to argue for exceptions to human rights, many people rightly fear that even to think of human rights as ways of equitably promoting well-being would make them much less secure. I discuss the paradox of direct consequentialism at great length in chapter 5, in part to try to allay this concern, and then, in chapter 14, I respond to this objection directly. Though the main principle explains the exceptions to human rights norms, it does not support our using *it* as a ground-level principle to justify exceptions to human rights norms.

Conclusion

Mill made the first attempt at a meta-level consequentialist explanation of the moral appropriateness of autonomy rights. As I interpret him, the early metaphysical Rawls expanded the project to try to provide a meta-level consequentialist explanation of the moral appropriateness of both autonomy and democratic rights (i.e., liberal) rights. My goal is even more ambitious: to try to provide a meta-level consequentialist explanation of autonomy rights, democratic rights, and five other categories of rights, as well—economic rights, negative and positive opportunity rights, social insurance rights, and privacy rights. The project is to explain why robust, inalienable rights of all those kinds should be universally guaranteed to all normal human adults by governments everywhere—that is, to explain why these rights *should be* universal. For that reason, I think of them as the rights that should be recognized as *human rights*. My ultimate goal is to provide a consequentialist meta-theoretic explanation of the content of these human rights. This is the *consequentialist project for human rights*. No single book could complete the project, so my aspiration for this book is to contribute to the project and, thus, to make it more plausible that the project might be successfully completed.

Because the methodology for my contribution to the consequentialist project for human rights is largely bottom-up, I can undertake the project even though I have no definition of *well-being* and I have no formula for its equitable distribution. A meta-theory of human rights is a theory of a moving target. If such a theory were to provide a complete vindication of current opinions about what is just or about what human rights should be, it would be a failure, because there is nothing more certain in moral matters than that current opinions can be improved.

A meta-theory of human rights should provide guideposts for potential improvements in current opinions and provide the resources for understanding why future changes are improvements (when they are). This is a tall order for any theory to have to fill. And it is one that any theory is bound to come up short on. Any normative theory of justice or of human rights, including this one, is bound to be imperfect, and thus improvable. This potential for

improvements in our current opinions and in our normative theories is the basis for a dynamic between theory and practice in which, over time, a theory can help us to improve our ground-level moral judgments and our ground-level judgments can help us to improve the theory. My goal in this book is to contribute to that process.

It is important at the outset for me to address a potential misunderstanding. Some people think that a right cannot be a human right if there is reasonable disagreement about it. As I explain in chapter 8, I think this seriously misunderstands the historical-social process by which human rights have been and are being discovered. In any case, let me say right here that almost everything in this book is subject to reasonable disagreement.

Exceptions to Libertarian Natural Rights

In the previous chapter, I proposed that we pay attention to the historical process of making exceptions to ground-level primary moral norms and principles. In this chapter, I compress and idealize some of that history to briefly illustrate the bottom-up reasoning involved, as an example of what it is that the main principle is designed to explain. I have claimed that the main principle provides a sufficient condition for the moral appropriateness of changes in ground-level primary moral judgments in any tradition that has passed the consequentialist threshold. I illustrate this claim by considering the natural rights tradition that developed in the West, because of the great power of that tradition and because the best way of explaining my theory of human rights is as a development from that tradition. However, it is important to realize that the main principle transcends any particular moral tradition to apply to all moral traditions that have passed the consequentialist threshold.

In the following conversation, three philosophers attempt to formulate ground-level moral principles for the state of nature, a situation in which there are no governments and thus no legal obligations. The state of nature is a heuristic for thinking about moral obligation in a way that avoids confusing it with legal obligation. The state of nature can play this heuristic role without our being committed to thinking that any such state ever actually existed.

An Example of Changes in Ground-Level Moral Principles through Bottom-Up Reasoning

Three philosophers, Moses, Fred, and Bob, were discussing the state of nature. Moses asked them to consider the following example: Anne is sitting minding her own business. Adolph comes up to her, pulls out a gun, and threatens to kill her unless she will be his slave.

Moses, Fred, and Bob all agreed that it would be wrong for Adolph to coerce Anne in this way. Moses suggested the following ground-level principle to explain why it would be wrong:

The Simple Prohibition on Coercion. It is wrong to coerce another human being by threatening to kill her.