THE PROBLEM OF PUNISHMENT

DAVID BOONIN

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521883160

The Problem of Punishment

In this book, David Boonin examines the problem of punishment, particularly the problem of explaining why it is morally permissible for the state to treat those who break the law differently from those who do not. Boonin argues that there is no satisfactory solution to this problem and that the practice of legal punishment should therefore be abolished. Providing a detailed account of the nature of punishment and the problems that it generates, he offers a comprehensive and critical survey of the various solutions that have been offered to the problem and concludes by considering victim restitution as an alternative to punishment. Written in a clear and accessible style, *The Problem of Punishment* will be of interest both to anyone looking for a critical introduction to the subject and to anyone who is already familiar with it.

David Boonin is associate professor of philosophy at the University of Colorado, Boulder. He is the author of *A Defense of Abortion* and *Thomas Hobbes and the Science of Moral Virtue*, as well as numerous articles on a variety of topics in ethics and applied ethics.

The Problem of Punishment

DAVID BOONIN
University of Colorado

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York
www.cambridge.org
Information on this title: www.cambridge.org/9780521883160

First published in print format 2008

ISBN-13 978-0-511-38833-0 eBook (NetLibrary)

ISBN-13 978-0-521-88316-0 hardback

ISBN-13 978-0-521-70961-3 paperback

For Leah, Eli, and Sadie – my greatest rewards

Contents

Preface *page ix*

1 The Problem of Punishment 1
1.0 Overview 1
1.1 What Punishment Is 3
1.2 What the Problem of Punishment Is 28

2 The Consequentialist Solution 37
2.0 Overview 37
2.1 The Act-Utilitarian Version 39
2.2 The Rule-Utilitarian Version 62
2.3 Other Utilitarian Versions 77
2.4 Nonutilitarian Versions 79

3 The Retributivist Solution 85
3.0 Overview 85
3.1 Desert-Based Retributivism 87
3.2 Forfeiture-Based Retributivism 103
3.3 Fairness-Based Retributivism 119
3.4 Other Versions of Retributivism 143

4 Other Solutions 155
4.0 Overview 155
4.1 The Consent Solution 156
4.2 The Reprobative Solution 171
4.3 The Moral Education Solution 180
4.4 The Self-Defense Solution 192
4.5 Hybrid Solutions 207

5 The Appeal to Necessity 213
5.0 Overview 213
5.1 The Theory of Pure Restitution as a Response to the Appeal to Necessity 218

5.2 *Clarifying the Theory of Pure Restitution* 220
5.3 *The Harm to Society Objection* 224
5.4 *The Irreparable Harms Objection* 235
5.5 *The Victimless Wrongdoing Objection* 245
5.6 *The Failed Attempts Objection* 249
5.7 *The Nonharmful Endangerment Objection* 255
5.8 *The Mitigating Excuses Objection* 256
5.9 *The Rich Offender Objection* 259
5.10 *The Poor Offender Objection* 261
5.11 *The Insufficient Deterrence Objection* 264
5.12 *The Insufficient Reprobation Objection* 267
5.13 *The Punishment as Restitution Objection* 269

Bibliography 277
Index 293

Preface

Most of my beliefs about relatively uncontroversial moral matters are relatively uncontroversial. We should generally be nice to each other, keep our promises, tell the truth, refrain from committing theft, arson, murder, and so on. Most people believe these things, and I do, too. Some of my beliefs about moral matters, of course, are more controversial. But these tend to be beliefs about matters that are themselves more controversial, things like abortion, animal rights, cloning, and so forth. If there were an uncontroversial position on these issues, chances are good that that position would be mine as well.

As far as I can tell, in fact, there is just one conspicuous exception to this general pattern. Most people believe that if it is just and reasonable for the state to prohibit a given form of behavior, then it is morally permissible for the state to punish those who persist in engaging in it. I don't believe this. I don't believe that it is morally permissible for the state to punish people for breaking the law. And I don't believe this because belief in the moral permissibility of legal punishment strikes me as inconsistent with many other things that I do believe.

I've felt this way about punishment for quite some time, and this fact has always struck me as puzzling. If most of my moral beliefs are the same as the moral beliefs of most other people, and if my rejection of the moral permissibility of punishment seems to be the natural upshot of most of my moral beliefs, then shouldn't most other people reject punishment too? Have most other people recognized something important about punishment that I've failed to see? Or have I been struck by something important about punishment that most people have overlooked? I have wondered about these questions for a long time.

The best way I know to learn about a philosophical problem is to teach a course on it. And so, several years ago, knowing virtually nothing about the philosophical literature on the subject, I designed and started to teach a course on the problem of punishment. The result of that undertaking is

this book. I hope it will be of use to those who are looking at the literature on the subject for the first time as well as of interest to those who are already deeply familiar with it.

Since this book could not have been written without the thoughtful and enthusiastic questions and comments I received from my students when I taught this material over the past several years, my single greatest debt is to them and I am pleased to acknowledge their contributions first. In addition, I received a great deal of feedback from a number of my colleagues, both in the form of comments on earlier drafts of parts of my manuscript and during discussions at a series of talks I gave over a period of a few years at our department's Center for Values and Social Policy. As far as I can tell, I did not succeed in convincing any of my colleagues to adopt my views, but many of them certainly succeeded in convincing me of numerous respects in which my work in progress needed improvement. Of these, I particularly thank Claudia Mills and my former colleague Jim Nickel for being so generous with their valuable time. For comments on extensive portions of earlier versions of the manuscript, I am also grateful to my father, Leonard Boonin, a former student of mine, Dan Korman, Thom Brooks, Jason Hanna, Jason Wyckoff, several anonymous referees for Cambridge University Press, and, above all, Gerald Postema, whose extremely sharp and focused comments on the first version of the manuscript that I submitted to Cambridge were indispensable in helping me to revise it with an eye toward publication.

For financial assistance in helping me to complete this project, I am grateful to the National Endowment for the Humanities for a summer research fellowship that I used to draft an early version of what eventually became Chapter 5, to the University of Colorado for granting me a research leave that I used to begin work on the project and a sabbatical that I later used to complete a full draft of the manuscript, and to the University of Canterbury for an Erskine Fellowship that enabled me to spend that sabbatical working on the book while in residence at its School of Philosophy and Religious Studies. Getting away from it all to the South Island of New Zealand was the perfect way to clear my mind and make the final push toward completing this project, and I'm grateful to everyone there who made my family's stay so memorable, especially Philip Catton and Carolyn Mason.

Finally, I am happy to acknowledge my immeasurable gratitude to my wife, Leah, and to my children, Eli and Sadie. This book is for them.

1

The Problem of Punishment

1.0 OVERVIEW

Legal punishment involves treating those who break the law in ways that it would be wrong to treat those who do not. Even if we assume that those who break the law are responsible for their actions and that the laws they break are just and reasonable, this practice raises a moral problem. How can the fact that a person has broken a just and reasonable law render it morally permissible for the state to treat him in ways that would otherwise be impermissible? How can the line between those who break such laws and those who do not be morally relevant in the way that the practice of punishment requires it to be? This is the problem of punishment.

The problem of punishment has generated a large and increasingly sophisticated literature with a wide variety of attempted solutions. This book contributes to that literature in two ways. First, it offers a comprehensive, up-to-date introduction to the contemporary literature by providing a detailed account of the nature of punishment and of the problem it poses, followed by a survey of the many solutions to the problem in the current literature. Second, it provides a critical evaluation of these solutions both as a means of introducing the reader to the various debates that these solutions have generated and as a way of defending a particular thesis about the problem that stands in stark contrast to the position taken in the vast majority of the literature on the subject. This is the thesis that there is no solution to the problem of punishment and that it is morally impermissible for the state to punish people for breaking the law.

The claim that it is morally impermissible for the state to punish people for breaking the law is likely to strike most people as implausible, if not absurd. While debates persist about precisely which forms of behavior a government may justly and reasonably prohibit, there is widespread agreement that if it is appropriate for a state to prohibit a particular form of behavior, then it is permissible for the state to punish

those who engage in it. The thesis defended by this book in response to the literature on the problem of punishment will likely be met with a good deal of resistance.

Much of this resistance is likely to be based on the belief, or at least the assumption, that there is a satisfactory defense of the moral permissibility of punishing people for breaking the law. Philosophers and legal theorists have typically sought to justify this practice either by appealing to the consequentialist claim that the presumed benefits of punishment are sufficient to render it morally permissible or by relying on the retributivist claim that punishment is justified because it is a fitting response to wrongdoing, regardless of its consequences. Others have attempted to justify legal punishment on a variety of additional grounds, such as the claim that such punishment is a form of moral education, social expression, or collective self-defense. Still others have appealed to some combination of these views. If resistance to this book's thesis rests on the belief that one or more of these attempts to establish the moral permissibility of legal punishment is successful, the only way to try to overcome this resistance is to try to demonstrate that all of these attempts are unsuccessful. That is the task of the central chapters of this book. In Chapter 2, I explain and argue against a variety of consequentialist attempts to justify the claim that legal punishment is morally permissible. In Chapter 3, I explain and argue against a variety of attempts to justify this claim along retributivist lines. And in Chapter 4, I explain and argue against a variety of further attempts to support the moral permissibility of punishment that do not fall readily under either of these two headings or that attempt to fall simultaneously under both.

A second likely source of resistance to this book's thesis is the belief that there is no way for us to do without punishment. Punishment, on this understanding, is necessary, either as a condition for the existence of a social order at all or as a condition for the kind of social order that makes possible just relationships among its members. On either version of this appeal to necessity, the practice of punishing people for breaking the law is said to be necessary, and if a practice is necessary, then an argument against its permissibility may seem pointless at best, incoherent at worst. I respond to this second source of resistance to this book's thesis in the final chapter by presenting and defending a single counterexample to the claim made by the appeal to necessity. This is the proposal, most widely associated with Randy Barnett's provocative article "Restitution: A New Paradigm of Criminal Justice" (1977), that we do without punishment by embracing a system of compulsory victim restitution. Following Barnett, I refer to this proposal as the "theory of pure restitution," and I argue that the theory is good enough to warrant rejecting the appeal to necessity. I do not insist that compulsory victim restitution is the only acceptable alternative to punishment or even that it is the best alternative. But I do

argue that it is an acceptable alternative. If it is an acceptable alternative, then punishment is not necessary. And if punishment is not necessary, then the appeal to necessity fails to undermine this book's central claim that it is morally impermissible for the state to punish people for breaking the law.

Finally, it seems likely that at least part of the resistance to this book's thesis lies in the failure to recognize that punishing people for breaking the law requires moral justification in the first place. The practice of punishment, after all, is ubiquitous. Ubiquitous practices are rarely called into question. Addressing this important concern is the goal of this introductory chapter. In Section 1.1, I explain why a critical assessment of punishment must begin with a definition of legal punishment (1.1.1), briefly present some criteria for adjudicating between rival definitions (1.1.2), and then present and defend a definition of legal punishment that does best by these criteria (1.1.3–1.1.7). I conclude this section by showing that, by this definition, punishment is importantly different from compulsory victim restitution (1.1.8). In Section 1.2, I explain why punishment, so understood, requires moral justification and poses a genuine moral problem (1.2.1), respond to two arguments against this claim (1.2.2–1.2.3), and conclude by identifying and explaining two tests that any solution to the problem of punishment must pass to be considered successful (1.2.4). This analysis, in turn, sets the stage for the remainder of the book, in which I argue that no solution to the problem of punishment passes both of these tests and that we should abolish our practice of punishing people for breaking the law.

1.1 WHAT PUNISHMENT IS

1.1.1 The Need for a Definition

When we talk about the moral permissibility of legal punishment, what, precisely, do we mean? A general answer to this question is easy: we mean such practices as the state's imposition of monetary fines, forced incarceration, bodily suffering, and – in extreme cases – death. A more specific answer is more difficult. Simply illustrating punishment,[1] even by appealing to clear paradigmatic examples, is not the same as defining it.

But is a more specific answer necessary for our purposes? It is tempting to suppose that it is not. As long as we all know what counts as examples of punishment, it might be said, we can move directly to the task of arguing about whether or not it is morally defensible. Indeed, one book on punishment begins by declining to offer a definition of the term for

[1] Unless otherwise noted, when I say "punishment" in this book, I mean "legal punishment."

precisely this reason: "one does not require a definition of 'punishment' in order to recognize clear cases of punishment's being imposed and to distinguish such cases from those in which individuals are treated in ways that, although similar to punishment in certain respects, are nevertheless something else entirely" [Montague (1995: 1)]. An "understanding" of punishment is certainly needed, Montague concedes, but one can understand punishment well enough without defining it.

While the reluctance to begin a discussion of punishment by developing a clear, specific definition is understandable, however, it is ultimately misguided. For a fully satisfactory inquiry into the moral permissibility of punishment, it is not enough to point to examples and say either that they are cases of punishment or that they are cases of something else. One must also be able to identify the properties that make them something else. If one cannot do this, then one cannot fully determine what, precisely, makes the permissibility of punishment problematic. More importantly, if one cannot do this, then one cannot satisfactorily determine whether or not a purported justification of punishment succeeds in justifying punishment or only in justifying something very much like it. Indeed, as we will see in Chapter 4, Montague's own attempt to defend punishment on grounds of social self-defense fails in part for precisely this reason.[2] Even if the argument from self-defense succeeds, I will argue, the practice that the argument would justify lacks two of the necessary characteristics that any satisfactory definition of punishment must include. Montague's failure to define punishment at the beginning of his book results in his failure to see that what he is defending at the end of his book is not exactly punishment.

Finally, and perhaps most importantly for the purposes of this book, we cannot fully disentangle the importantly related practices of punishment and compulsory victim restitution without understanding what makes some cases cases of punishment and others cases of something else. Such disentanglement is crucial to the project of this book: it is necessary to see precisely why rejecting the claim that punishment is morally permissible does not entail rejecting the claim that compulsory victim restitution is morally permissible. For all of these reasons, then, we must begin our investigation by clarifying what makes some forms of treatment cases of punishment and others cases of something else. And it is difficult to see how to do this without a definition.

1.1.2 The Criteria for a Definition

So, we want a definition of punishment. But we do not want just any definition. We want a good one. What would constitute a good definition of punishment? First, it must be *accurate*. It must provide us with a set of

[2] See Section 4.4.5.

necessary and sufficient conditions that clearly demarcates cases of punishment from cases of something else. The results produced by this demarcation must cohere sufficiently well with what we mean by punishment when we argue about it and must do so over a sufficiently wide range of cases. If it is clear that responding to an offender's behavior by fining him, beating him, or executing him do count as punishments, for example, and that responding to his offense by writing him a check, throwing him a parade, or giving him a free meal do not, then an adequate definition of punishment must account for these judgments.[3] If it is unclear or indeterminate whether or not responses such as voter disenfranchisement, supervised probation, public shaming, or certain forms of taxation should count as forms of punishment, then a good definition should help us to make sense of these facts as well.

Second, a good definition of punishment must be *illuminating*. A definition may be accurate, successfully discriminating between cases of punishment and cases of something else, but if it does so only because it contains various stipulations that are thrown in solely to produce the desired results and have no further independent motivation, then the definition will be unacceptably ad hoc. When we appeal to it in asking whether or not a particular act counts as an act of punishment, such a definition will give us the correct answer, but it will do nothing to demonstrate why the answer is correct. In part, we want a good definition to get at the essence of the thing defined, to tell us not just that a given subject belongs in a certain class with certain other subjects, but in virtue of what fact or set of facts this is so.

Finally, a good definition of punishment must be *neutral* on the question of whether or not punishment is morally permissible. A definition is unacceptable if it begs the question one way or the other, with respect to either the merits of punishment in general or the merits of any kind of justification of punishment in particular. If, for example, one attempted to discriminate between punishment and mere private vengeance by saying that punishment is "authorized" while private vengeance is not, and if part of what one meant by an act's being authorized was that it was legitimate, then the resulting definition of punishment would unacceptably beg the question in favor of the claim that punishment is morally permissible. If one defined punishment so that part of what made an act a punishment is that it was justified because of its effects

[3] The claim that a good definition must accurately capture actual usage of the term 'punishment' does not mean that in order to use such a definition, a defender of the permissibility of punishment must defend the permissibility of all forms of punishment. It means that a defender of punishment must acknowledge that capital punishment and corporal punishment are forms of punishment, for example, but it does not mean that he or she must insist that they are morally permissible.

on society, or that it was not justified in this way, then the result would fail to be neutral with respect to the various competing solutions to the problem of punishment. In short, we want a definition of legal punishment that respects and reflects both our beliefs about what counts as punishment and our puzzlement over what, if anything, renders it morally permissible for the state to punish people.

1.1.3 Harm

A definition that satisfies these requirements can be obtained by testing various conditions against our intuitive reactions to clear, paradigmatic instances of legal punishment. As already noted, such cases include monetary fines, forced incarceration, bodily suffering, and, in extreme cases, death. So, we should begin by asking what these various practices have in common.

Perhaps the most obvious quality that these practices have in common is that they are all in some way bad for the person on whom they are inflicted.[4] This point is often expressed by saying that punishment necessarily involves "pain," but this way of putting things is unsatisfactory.[5] A murderer, for example, could be executed painlessly, and this would clearly be bad for him even if he does not experience pain. The same problem arises if punishment is defined, as it sometimes is, in terms of subjecting people to experiences that are "unpleasant." Other writers have attempted to capture the sense in which punishment involves something negative for the person on the receiving end by saying that punishment involves an "evil," but this runs the risk of defining punishment as something that is, at least in itself, a wrong; and this, in turn, would violate the requirement of neutrality by begging the question against those retributivists who maintain that the treatment that punishment inflicts on an offender is not merely allowable but a positive good. Finally, some writers have defined the negative effect of punishment on the person who is punished in terms of the language of rights. Punishment, on this account, involves depriving someone of what would otherwise be a right. If one holds the view that losing a right is always bad for someone, then putting things in terms of rights poses no real difficulties for an analysis of punishment as something that is bad for someone. But if, as seems plausible to me, there can be cases in which a person loses a

[4] That punishment involves treatment that is, by some measure, of negative value for its recipient is accepted by virtually every philosopher who has written on the subject, including such historical figures as Plato, Aristotle, Aquinas, Hobbes, Locke, Kant, and Hegel, as well as more recent writers such as Flew, Benn, Hart, McCloskey, Honderich, and Primoratz [for citations, see Adler (1991: 285–6)].

[5] See, e.g., Newman, who insists that "Punishment must, above all else, be painful" (1983: 6), and Corlett (2001: 68).

right but is not made worse off by this loss, then such cases would seem to provide a good reason not to link punishment to rights by definition. A woman who is physically incapable of becoming pregnant, for example, might still have a legal right to an abortion, and if depriving her of that right would in no way be bad for her, it is difficult to see how it could count as punishing her. It therefore seems more sensible to say that acts of punishment all, in some way, make the person who is punished worse off than she would otherwise be. If an offender received a monetary prize for her offense, or a paid vacation, a relaxing massage or life-extending therapy, for example, we would not be inclined to say that she had been punished for her transgression. And so, a natural starting point in generating a definition of punishment is to say that punishment *harms* the person who is punished, where harming someone means making her worse off in some way, which includes inflicting something bad on her or depriving her of something good. I will refer to this as the "harm requirement."

1.1.3.1 *The Beneficial Consequences Objection*

A critic of the harm requirement might object that this requirement neglects the beneficial long-term consequences that punishment can have for the person who is punished. Adler, for example, who rejects the claim that harmfulness is an essential property of punishment, appeals to what he calls the "conscientious punishee," the offender "who wants to submit to punishment, who believes that she can achieve reconciliation, atonement, expiation, renewed innocence, greater moral knowledge, or some other good by undergoing the punishment" (1991: 91). Indeed, as we will see in Section 4.3, a number of writers have claimed not only that punishment ultimately benefits the offender who is punished, but that the moral permissibility of punishment is grounded in this very fact. A definition of punishment that incorporates the harm requirement would therefore seem to beg the question against such a position, ruling out the possibility that punishment might be justified as ultimately good for the person punished by definitional fiat. This, in turn, would violate the neutrality requirement established earlier, rendering the definition unacceptable.

This objection to the harm requirement is understandable, but it is also mistaken. The harm requirement maintains that for a certain treatment to count as a punishment, it must harm the recipient. But it is neutral on the further question of whether or not being subject to such a harm might produce beneficial consequences in the future, including beneficial consequences that are great enough to outweigh (and perhaps even to justify) the immediate harmful ones. Consider, for example, a child who is spanked as a (nonlegal) punishment for having hit another child. The parent who punishes a child in this way may believe that spanking will

make him understand more fully why what he did was wrong, and that this, in turn, will contribute to the child's moral development in various important ways. If this is so, then spanking the child now will ultimately benefit him in the future. But all of this is perfectly consistent with the harm requirement. Indeed, it presupposes it. For if spanking the child does benefit him in this way, then this will be so precisely because it involves inflicting a harmful treatment on the child as a means of demonstrating to him how it feels to be on the receiving end of such harmful treatment. If the spanking were not harmful to the child (if, for example, it felt just like being pleasingly caressed), then it would not have the desired educative effect of showing what it is like to be a victim of wrongful treatment in the first place. So, considerations of the possible long-term benefits of punishment provide no reason to reject the harm requirement. If anything, they provide further reason to accept it.[6]

1.1.3.2 The Masochist Objection

A second objection to the harm requirement is that it is subject to refutation by counterexample. Most people, for example, strongly dislike being physically beaten. But some people, apparently, do not. Most people would find incarceration highly unpleasant. But some people, perhaps, would not, and others, depending on their circumstances, might find it preferable to the available alternatives. And so, it might be urged, we can say at most that punishment involves treatments that are *typically* harmful or that are considered undesirable by *most* people, but we cannot say that this is so of punishment in every instance. You and I might strongly prefer not to be whipped, for example, and so this punishment would be harmful to us, but a masochist might enjoy a beating; and, if he did, it would remain a form of corporal punishment nonetheless. Since such cases apparently involve acts that are acts of punishment but that do not harm their recipients, they seem to demonstrate that the harm requirement is not accurate over an important (even if somewhat limited) range of cases.[7]

The objection that appeals to cases such as the masochist rests on two claims: that in such cases the treatment in question does not harm the recipient and that it counts as punishment nonetheless. A defender of the

[6] It is possible, of course, that a proponent of the objection might insist that the benefits of submitting to punishment are immediate rather than delayed. But if the recipient of a given treatment is benefited at the moment that the treatment begins, it is not clear what reason we would have for considering it to be a punishment in the first place. If a pleasant caress on the child's back benefits him immediately and also somehow teaches him that it is wrong to hit other children, for example, then it may well serve the same purpose as a spanking for educative purposes, but it would clearly fail to count as punishment and so would again fail to provide a counterexample to the harm requirement.

[7] This problem is raised by, e.g., Kasachkoff (1973: 364–5) and Snook (1983: 131).

harm requirement might reject the objection's first claim and argue that even if the masochist enjoys being beaten, a beating is still something that is objectively harmful to him. Similarly, even if a homeless or insecure person prefers the security of prison to the unpredictability of life on the outside, one could argue that the restriction on his freedom of movement is objectively a grave harm to him even if he doesn't particularly mind it.

But even if the objection's first claim can be sustained in a significant range of cases, the second should be rejected outright. For if we concede that the masochist is not harmed by being whipped or that the homeless person is not harmed by being imprisoned, then we have two good independent reasons to conclude that he is not punished either. And if he is not punished, of course, then even if he is not being harmed, he cannot serve as a counterexample to the claim that punishment requires harm.

The first reason to believe that these attempted counterexamples fail in this way arises because there is a conceptual symmetry between punishment and reward. What is true of punishment in one direction, that is, must be true of reward in the other. Yet, in the case of reward, it should be clear that a person has not been rewarded for doing a good deed if the treatment that she receives in response does not in fact end up benefiting her. Suppose, for example, that I give you a piece of candy because you did me a favor last week, but the candy causes a severe allergic reaction. We might say that I tried to reward you for your good deed or that I intended to reward you, but we would not say that you had, in fact, been rewarded. And we would not say this precisely because you had not been benefited. Since it seems reasonable to presume that reward and punishment are symmetrical in this respect, this provides support for the claim that the offender who is not actually harmed by the treatment he or she receives is not actually punished by it.

The second reason to believe that without real harm there is no real punishment arises from cases in which we believe that no harm is done because of some particular fact about the treatment itself. When a stay in a minimum-security prison for white-collar criminals seems to resemble nothing more than an all-expenses-paid vacation at a comfortable resort, for example, people do not consider the offender to have been punished and they complain about his being treated so leniently for precisely this reason.[8] Our intuitive response to punishments that seem clearly nonharmful and to attempts to reward that clearly do not benefit both vindicate the claim that the harm requirement is a core component of our concept of punishment. And so, the apparent counterexamples to the

[8] When it was reported that the son of former vice presidential nominee Geraldine Ferraro was serving his sentence for a drug conviction in a $1,500-a-month luxury apartment, for example, the public outcry over the case prompted the governor of Vermont to discontinue the house arrest option for drug offenders [Tunick (1992: 3)].

harm requirement, in which it seems that a person is punished but is not harmed, in the end do not undermine the harm requirement but once again reinforce it.

I think that these considerations suffice to defend the harm requirement from what might be called the "masochist objection," but there is one more concern that might be raised at this point. For if we agree that the masochist who is not harmed by his whipping is not punished by it either, it can seem that we must therefore conclude that whipping is not a form of punishment after all. And that result can seem sufficiently counterintuitive to force us back to the conclusion that the masochist really is being punished and that punishment therefore really does not require harm. This worry about my rejection of the masochist objection is understandable, but it is ultimately misguided. The reason is that there is a crucial difference between saying that a particular person has been subjected to a form of treatment that is a form of punishment and saying that this person has, in fact, been punished. And even if it is possible that some people are not harmed by being subjected to forms of treatments that are uncontroversially characterized as forms of punishments, this does not mean that we must say that such people are actually punished by such treatments.

Since this response to the objection may at first seem puzzling, an analogy may be of use. Consider a doctor who administers a sedative to a patient. An essential property of a sedative is that it makes people sleepy. But just as there are some people who may be delighted by some forms of punishment, there may be some people who are stimulated by some forms of sedatives. If the doctor gives such a drug to such a patient, then what she gives the patient might still be properly characterized as a sedative because of its general properties, but this does not mean that in giving the sedative to this particular patient she actually sedates the patient. Similarly, if the state inflicts a form of corporal punishment on someone who is not harmed by it, then while it may be proper to continue to refer to this treatment as a form of punishment (since it is a form of treatment that does, in general, harm people), this does not mean that in administering it to this particular offender the state will in fact be punishing him. It will, at most, be attempting to punish him.[9]

[9] And in at least some cases, it will not even be clear that it should be considered an attempt at punishment, let alone a successful attempt. After his lawyer reached a plea bargain agreement with Oklahoma City prosecutors for a thirty-year prison sentence for two charges of shooting with intent to kill and one weapons violation, for example, Eric James Torpy insisted that he would rather get thirty three years to match the uniform number of his basketball hero, Larry Bird. The judge in the case was quoted as saying that "We accommodated his request and he was just as happy as he could be" ("Man Asks for More Jail Time to Honor Bird" 2005). Although three extra years in prison would generally be considered a form of (additional) punishment, it is difficult to believe that the judge

1.1.3.3 The Community Service Objection

A final objection to the harm requirement also turns on the claim that it is subject to refutation by counterexample. While the masochist objection focuses on anomalous people who seem not to be harmed by treatments that would harm most of us, this objection focuses on a somewhat anomalous punishment that seems not to harm most people. In particular, the objection maintains that while the harm requirement may be consistent with several of the most commonly recognized forms of punishment, it fails to account for cases of a less standard but still uncontroversially punitive treatment: cases in which an offender is sentenced to perform community service.[10] Adler, in particular, has argued that, at least for offenders who want to accept their punishment because they believe they will benefit from it in the long run, mandatory community service is a genuine form of punishment but is not harmful in any significant way to the offender (1991: 91–2). Adler points out that community service can include many behaviors that are not undesirable or unpleasant, such as coaching a sports team or working with handicapped people, and cites a study showing that many offenders continue to volunteer for such projects after their sentences have been completed. If mandatory community service can be both punitive and nonharmful, then the harm requirement must again be rejected.

But it is the community service objection itself that must be rejected. This objection to the harm requirement fails to take into account the difference between, say, coaching a soccer team made up of disadvantaged children and being *compelled* to coach such a team.[11] If I have always wanted to coach such a team and you give me the opportunity to do so, then you benefit me. If I would prefer to do something else and you

thought he was punishing Torpy further by giving him the extra three years that would make him happy, let alone that doing so did, in fact, punish him.

[10] The case of an offender who is sentenced to probation might also be raised as a possible counterexample to the harm requirement. Probation is a relatively common sentence and, at least in typical cases, if the offender does not violate any of the terms of his probation, he will not suffer any harmful consequences as a result of his offense. But being subject to these requirements seems plausibly to count as a harm in itself (being deprived of the liberty to drink alcohol, to visit certain people or certain areas, and so on), and if one views the terms of probation as in no way harmful, it seems likely that one will, for that very reason, view a sentence of probation as something like a suspended sentence: not as a form of punishment but as an alternative to it. On either account, then, it seems implausible to construe probation as a genuine counterexample to the harm requirement.

[11] In a different context, Hampton notes that in the case of mandatory public service, "what makes any experience the suffering of punishment is not the objective painfulness of the experience, but the fact that it is one the wrongdoer is *made* to suffer and one which represents his *submission* to the punisher" (1988c: 126). Duff also notes this weakness of the community service objection in his review of Adler's book (1993: 181).

coerce me into coaching the team, then you harm me even if I end up enjoying myself and want to continue coaching the team after my sentence has been served. An offender who is forced to do something she would otherwise not do is thereby harmed; for this reason, such offenders fail to serve as counterexamples to the harm requirement. If an offender who has always wanted to coach such a team is required to do so, then it may well be true that he is not harmed. But for the reasons given in the previous section, it would also seem right to conclude that he is not thereby punished.

1.1.4 Intentional Harm

I have argued thus far in defense of the harm requirement. If subjecting a particular offender to a particular treatment does not harm her, then even if the treatment is, in general, a form of punishment, she has not been punished. The harm requirement accurately captures part of what is distinctive about punishment. It helps, for example, to distinguish correctly cases of punishment from cases of reward. But the harm requirement alone is not enough. For there are practices that involve inflicting the same kinds of harm that are inflicted in cases of punishment but that are clearly not cases of punishment. Consider, for example, the following two pairs of cases:

1. Larry marries Laverne and is charged a fee for processing his marriage license.
2. Moe marries both Betty and Veronica and is charged a fine for violating antipolygamy laws.

3. Curly is found not guilty of murder by reason of insanity and is confined against his will for the rest of his life.
4. Shemp is found guilty of murder and is confined against his will for the rest of his life.

In all four cases, a person does an action and is then harmed by the state as a result. Furthermore, the kind of harm that is incurred in (1) is the same as the kind incurred in (2), and the kind incurred in (3) is the same as the kind incurred in (4). But while (2) and (4) are clearly cases of punishment, (1) and (3) are clearly not.

Several considerations are required to account fully for the difference between (2) and (4), on the one hand, and (1) and (3), on the other. For the purposes of this section, however, we can focus on one feature: that which arises from the distinction between intentionally causing a harmful effect and foreseeably causing a harmful effect. Consider a patient who has been diagnosed with cancer and encouraged to undergo chemotherapy. She is told that the chemotherapy will have two effects: it will kill the cancer cells and it will cause hair loss. When this patient agrees to the

procedure, she does so with the intention of killing the cancer cells. She foresees that this will also cause hair loss, but this is not her intention. This can be cashed out in counterfactual terms. If the chemotherapy would kill the cancer cells but not cause hair loss, she would still undergo it. If it would cause hair loss but not kill the cancer cells, she would not. Because of these facts, we can say that she intends to kill the cancer cells but merely foresees that she will loss her hair.

This same distinction can be applied to the question of the role that inflicting harm plays in the institution of punishment. When a person is found not guilty of murder by reason of insanity, the state may determine that to protect the public, he must be locked up in a mental institution. In doing so, the state recognizes that its action will seriously harm the person,[12] but harming him is not its intention. Its intention is merely to protect the public, and it would lock him up even if this did not harm him. Similarly, when the state charges a fee for processing a marriage license, it understands that the cost imposes a harm on those getting married, but this is not its intention. Its intention is merely to recover the costs involved in processing the relevant paperwork, and it would charge the same fee even if, for some reason, couples getting married benefited from paying it.[13] When the state punishes someone, on the other hand, it inflicts various harmful treatments on him *in order* to harm him. It is not merely that in sentencing a prisoner to hard labor, for example, we foresee that he will suffer. Rather, a prisoner who is sentenced to hard labor is sentenced to hard labor *so that* he will suffer, and if a given form of labor turned out to be too pleasant and enjoyable, he would be sentenced to some other form of labor for precisely that reason. As Benn puts it, the "unpleasantness" of punishment is not merely an incidental byproduct or side effect of it, but rather is "essential" to punishment (1967: 8).[14]

[12] In some cases of this sort, of course, the state might determine that the person in question is also a threat to himself. In these cases, the person might be helped rather than harmed by being involuntarily confined. For purposes of this example, however, we can simply stipulate that we are talking about cases in which the person is clearly able to take care of himself, but poses an unacceptable risk to others.

[13] Regulations that require people to compensate the government for the costs their actions impose on it are widespread and are clearly understood to be nonpunitive. A rule approved by the Boulder County Board of Commissioners in Colorado in 2005, for example, requires that organizers of public protests that take place on public land pay the county to cover the costs of cleaning up after the event, and it is uncontroversial that this measure does not amount to punishing people for exercising their right to freedom of assembly (Miller 2005: 3A).

[14] Since the claim that punishment involves not merely harm but intentional harm is crucial to much of the argumentation in this book, it may be worth noting that the claim that the negative consequence for the offender is brought about intentionally is almost universally accepted in the literature on punishment. An extremely small sample of those who explicitly endorse it, selected more or less randomly from the literature I have examined

This is not to insist, it is important to emphasize, that the offender's suffering must be intended for its own sake. That would reduce punishment to sadism. Rather, it is to maintain that the punisher intends to harm the recipient of the punishment and does not merely foresee it, even if this harm is, in turn, intended for the sake of some further end. When a parent punishes her child in a nonlegal context by spanking him, for example, the pain inflicted on the child is not simply a foreseen side effect of the spanking, as it might be in the case of the pain caused by removing a splinter. If the splinter came out painlessly, the parent would not reinsert it in order to pull it out again in a more painful manner, but if the first spank was too mild to cause any pain, the parent would spank again, and harder. But while the parent who spanks her child thus clearly intends the pain that she causes and does not merely foresee it, she causes the pain not as an end in itself, but rather for the sake of some further end, such as educating the child or deterring him from committing similar infractions in the future.[15] It might at first seem odd to think that whether or not an act is an act of punishment could depend on facts about the intentional states of the punisher, but on reflection it should seem clear that this must be so. If you see an adult hitting a child, for example, you cannot know if the adult is disciplining the child or simply attacking the child without knowing the reason. If you see a uniformed official forcing a laborer to lift a heavy rock, you cannot know whether what you see is a prisoner being punished or a slave being exploited without knowing why the laborer is being forced to lift the rock.[16] And so, if you see an offender being subjected to a harmful treatment, you cannot know

in preparing this book, includes Ducasse (1968a: 9; 1968b: 34–5), Kasachkoff (1973: 367), Bean (1981: 2), Primorac [sic] (1981: 205), Burgh (1982: 193), Sverdlik (1988: 190, 198), Stephenson (1990: 229), Nino (1991: 258), Fatic (1995: 197), Wright (1996: 27), Clark (1997: 25), Scheid (1997: 441), Corlett (2001: 68), Gert, Radzik, and Hand (2004: 79), and Golash (2005: 1, 45).

[15] That the parent intends the harm only as a means and not as an end, it may be worth pointing out, does not undermine the counterfactual analysis of intentions presented here. A critic might point out that there is a counterfactual situation in which the parent would not harm the child: if the parent could achieve her ultimate end of educating or deterring the child without harming him, after all, surely she would. But this fact does not pose a problem for the claim that there is an important difference between the spanking case and the splinter-removing case. What matters is that in the spanking case but not in the splinter case, the parent has, in fact, chosen to use pain as a means to achieve her end, even if, in both cases, she would prefer not to.

[16] For a particularly poignant illustration of this phenomenon, see the nuanced discussion of the Soviet gulag system in Applebaum (2003). As Applebaum notes, whether or not a particular camp is best understood as a labor camp or a "punishment camp" depends largely on whether the prisoners were forced to work in order to produce needed goods or in order to make them suffer, and it is precisely in those cases in which the point of the suffering was not clear that it is unclear whether the system involved punishment or something else (221).

whether or not he is being punished without knowing why the offender is being so treated.[17]

As a result of these considerations, we must accept what I will call the "intending harm requirement": for an act to be a punishment, it must be done with the intent of harming the person being punished. Accepting this requirement is necessary to illuminate fully the difference between punishment, on the one hand, and such practices as charging user fees and requiring pretrial detention, on the other. Many examples can be used to support this basic point. Governments quarantine some of their citizens; call some of them for jury duty; conscript some of them into the military; subject some of them to curfews, taxation, and zoning regulations; build new highways or airports that adversely affect the quality of life of some of them; and so on. In all of these instances, and many more, the state acts in the full understanding that the act will harm some of its citizens. Clearly, though, none of these cases involves the state's punishing some its citizens. And part of what explains this is that the harm in these cases is foreseen but not intended.[18] The intending harm requirement, therefore, is needed to produce a fully illuminating definition of punishment.

The intending harm requirement is useful in other ways as well. First, it helps us to see more clearly not just what makes punishment distinctive,

[17] In some cases, there may be no straightforward answer to this question and thus no straightforward answer to the question of whether the act in question should be construed as punishment. In virtually every state in the United States, for example, prison inmates are denied the right to vote, and in several states, some or all of them are permanently barred from voting even after they have served their sentences. The justifications for felony disenfranchisement are murky at best, so accepting the intentional harm requirement will render it unclear at best whether this should count as a punishment. Since it seems to me unclear whether or not felony disenfranchisement should be understood as a punishment, this implication of the intentional harm requirement seems to count as a virtue rather than a vice [for a useful critical discussion of the practice, and of various attempts to justify it, see Pettus (2005: esp. chap. 4)].

[18] A further useful example concerns the so-called Megan's law provisions, which authorize states to post photographs of convicted sex offenders and other information about them on the Internet after they have been released from prison. Such provisions have been challenged as an unconstitutional extra punishment on the grounds that it harms them for their offense above and beyond the harm they received from serving their sentences. But in 2003, the U.S. Supreme Court ruled that although such provisions may harm them, they do not punish them, and in doing so it relied precisely on the distinction between foreseeable and intentional harm. As Justice Anthony Kennedy wrote for the majority in the 6–3 decision, "The publicity may cause adverse consequences for the convicted defendants, running from mild personal embarrassment to social ostracism," but the intention of the laws is "to inform the public for its own safety, not to humiliate the offender" (Holland 2003: 7A). Relatedly, the Court ruled that in determining whether or not pretrial detention constitutes a form of punishment, "a court must decide whether the disability is imposed for the purpose of punishment or whether it is but an incident of some other legitimate governmental purpose" [*Bell v. Wolfish*, quoted by Tunick (1992: 126)].

but what makes it distinctively problematic. For the difference between intending and foreseeing is not just useful in understanding what punishment is. It is also morally relevant in determining whether or not it is justified. Consider, to take a common example, a pair of cases involving a bomber who, during a justified war, drops a bomb on an enemy ammunitions plant surrounded by civilian housing. In the first case, the bomber drops the bomb in order to destroy the ammunitions plant, thinking that this will induce the enemy to surrender, while foreseeing that as a regrettable consequence of his act, some innocent civilians will be killed. In the second case, the bomber drops the bomb precisely to use the ammunitions plant as a means to kill innocent civilians, thinking that this will induce the enemy to surrender. Even though both bombers do the same act, with the same result, and with the same ultimate goal of inducing the enemy to surrender, it seems clear that what the second bomber does is worse than what the first bomber does. And it seems equally clear that this is so because the second bomber intentionally harms the civilians to achieve his purpose, while the first does not. Yet, as I have noted, intentional harm is precisely what the state inflicts in the case of legal punishment. It is one thing to justify the claim that it is morally permissible for the state to act in various ways while foreseeing that so acting will cause some of its citizens to suffer (e.g., changing the speed limit, modifying air pollution standards, imposing new regulations, raising taxes, or conscripting soldiers, all of which foreseeably cause harm to a significant number of people). It is quite another to justify the claim that it is morally permissible for the state to act in various ways *in order that* some of its citizens will suffer. Yet, this is precisely what must be justified in order to justify punishment. The intentional harm requirement, then, helps to illuminate more clearly what makes punishment distinctively problematic.

Finally, accepting the intentional harm requirement permits the discussion of the permissibility of punishment to proceed even if part of my earlier discussion of the nature of punishment is rejected. In Section 1.1.3.2, I responded to the masochist objection by maintaining that if a person is not harmed by a given act, then he is not punished by it either. But let us now suppose that I have been mistaken. For the purposes of the arguments to be developed in this book, this particular mistake can be ignored. For there is a final response that many people have offered to the various purported counterexamples to the harm requirement considered in the previous sections: we can say that even if, for example, the masochist is not harmed by the whipping he receives, the whipping is a punishment because it is done with the *intention* of harming him.[19] This

[19] See, e.g., Kleinig (1973: 24): "while I do not think it to be a necessary characteristic of punishments that they be *experienced* as impositions, I would insist that they be *intended* as impositions." The same view is taken by Newman (1978: 8) and Walker (1993: 1).

response does not, strictly speaking, rescue the harm requirement from its critics. Rather, it replaces the harm requirement with the intentional harm requirement. For the reasons given in the previous section, I do not believe that this substitution is necessary. But since many readers may disagree with this assessment, and since I believe that the problem with punishment can be treated just as effectively if the substitution is accepted, I will go ahead and accept it, at least for the sake of the argument. The moral problem of punishment, after all, remains just as vivid even if this emendation in the definition is accepted: it is just as difficult to justify the state's acting with the intent of harming some of its citizens, regardless of whether or not it succeeds in harming them.

1.1.5 Intentional Retributive Harm

So punishment involves, at the least, intentional harm. But it also involves more than this. If I punch you in the nose to make you suffer, then I harm you, and do so intentionally, but I do not punish you. This is because I am not harming you in response to some transgression of yours. So, an additional requirement is this: to be a punishment, an act must involve intentionally harming someone *because* he previously did a prohibited act. And since we are concerned in this book with legal punishment in particular, we can be more specific: to be a legal punishment, an act must involve intentionally harming someone because he previously did a legally prohibited act, which means that he is responsible for having done the act and that he had no valid legal excuse for doing so.[20] Call this the

[20] Sverdlik, in particular, points out that many definitions of punishment overlook this last requirement (1988: 183). To say that legal punishment is for a "legally prohibited act" need not beg any questions about whether acts of retaliatory warfare might be justified as punishment: even if they do count as punishment, they do not count as legal punishment [for a useful discussion of punitive justifications for acts of war, see Kemp (1996)]. And to say that punishment is for "a" legally prohibited act need not beg any questions about the possible relevance of prior convictions in determining the appropriate amount of punishment in any particular instance. On some views, for example, when a repeat offender is punished more severely for his third offense than he would be for his first, the extra severity of the punishment amounts not to punishing him for the third offense, but rather to punishing him again for his previous offenses; in this case, it would seem that he is being punished in response to more than one prohibited act. But if this analysis is be correct, we can simply construe the punishment itself as a combination of some punishment for one act and some punishment for another. To avoid any possible complications, however, throughout this book I will focus on cases involving first-time offenders. This restriction, moreover, begs no questions against the various attempts to justify punishment. If there is no justification for punishing someone the first time, after all, then there can be no justification for punishing someone the first time and then punishing him even more at a later time [for a useful critical discussion of the possible moral relevance of prior convictions, see Bagaric (2000)].

"retributive requirement."[21] The retributive requirement is needed to distinguish cases of punishment from cases of mere gratuitous injury. It therefore seems plainly well motivated. Is there any reason to reject it?[22]

1.1.5.1 The Mistaken Verdict Objection

One reason to reject the retributive requirement might seem to arise from the following simple fact: sometimes innocent people are mistakenly convicted of offenses they did not commit and serve sentences for those offenses. It thus seems clear that sometimes innocent people are punished. Yet, if the retributive requirement is correct, there can be no such thing as punishment of the innocent, since by definition punishment involves punishing someone for an offense she actually committed. Since there seem plainly to be cases in which the innocent are punished, and since the retributive requirement seems unable to account for this fact, it seems plain that the retributive requirement must be rejected.[23]

This reasoning, however, is misleading. This can perhaps be most clearly seen by again appealing to the structural symmetry between punishment and reward. Suppose that I have lost my beloved dog and have offered a $500 reward for her safe return. As you walk into my front yard for a visit one afternoon, she comes bounding up the steps, and I mistakenly conclude that you have found and returned her. So, I give you the $500. Clearly, what you have received from me is a benefit, and you have received it from me for just this reason. So, it is clearly true that you have been intentionally benefited by me. But even though this is true, it is not true that you have been rewarded by me, any more than it would be true that you have been repaid by me if I had given you the money under

[21] The retributive requirement is defended by a number of writers, including Flew (1954: 85), Quinton (1954: 58–9) and Schedler (1976).

[22] In addition to the reasons given later, one might object that the requirement makes no mention of a proportionate fit between the punishment and the offense. But this omission is perfectly appropriate. Executing someone for stealing a candy bar from a store is still a punishment, even if it is clearly an excessive punishment.

[23] This objection has been pressed by a number of writers, including Baier (1955: 130–2), Ewing (1963: 122), Locke (1963: 568–9), Gendin (1967–8: 235), Pratt (1968: 22), Scheid (1990: 456), and Hare (1986: 212). Kasachkoff (1973: 369) presents a closely related problem case: that of a teacher who punishes an entire class for an infraction committed by one unknown student. In this case, however, as in the case where one country takes military action against another in retaliation for an unprovoked attack, it should be clear upon reflection that the innocent students (or civilians) who suffer the consequences are not themselves being punished but rather are incurring the collateral damage produced by the punishment of the true offender. In addition, the classroom example can be rejected by appealing to the symmetry between desert and punishment appealed to in the text subsequent to this footnote: if every student in the class received a benefit in response to the act of one unknown Samaritan among them, it would be clear that the students who received the benefit but did not do the good act were not being rewarded but were simply benefiting from the good act of another.

the mistaken impression that I had borrowed it from you last week and now owed it to you. I may believe that I am rewarding you for something, I may intend the money to be a reward, but in fact what I am giving you is not a reward. And it is not a reward (or a repayment) precisely because you have done nothing to be rewarded (or repaid) for. The same is therefore true in the case of punishment. If you are mistakenly convicted of an offense you did not commit, then the judge may believe that she is punishing you and may throw you in jail with this intention, but the harm she imposes on you cannot, strictly speaking, be a punishment. An innocent person can suffer the harms caused by a particular punishment but cannot, strictly speaking, be punished. As Quinton, among others, has recognized, "The infliction of suffering on a person is only properly described as punishment if that person is guilty" (1954: 59).[24]

The claim that an innocent person cannot, strictly speaking, be punished seems to be correct. Indeed, it seems to be as clearly correct, and for the same basic reason, as is the claim that a person who has not loaned money to someone cannot, strictly speaking, be repaid. And if the claim is correct, then the mistaken verdict objection to the retributive requirement must clearly be rejected. But it is important to emphasize that the retributive requirement can be construed in a slightly different way so that it retains its basic force even if the objection is sustained.[25] In Section 1.1.4, I noted that the intentional harm requirement can be construed in two different ways: as the requirement that an act cause harm and that the harm be caused intentionally, or as the requirement that an act be intended to cause harm, regardless of whether it does so. And although I maintained that the former, stronger version of the requirement should be accepted, I noted that even in the latter, weaker version, it was sufficient to generate the problem of punishment, and so stipulated that I would accept the weaker version at least for the sake of the argument. The same can be done here with the retributive requirement. On a strong reading, the requirement is that to be a punishment, an act must be retributive. This would require punishment to involve harm that is both retributive and intentional. On a weak reading, the requirement is that to be a punishment, an act must be *intended* as retributive even if it turns out not to be because the recipient of the treatment is innocent. This would require punishment to be intentional retributive harm rather than retributive intentional harm. As for the harm requirement, I believe that the retributive requirement should be accepted in its strong version. But, again as for the harm requirement, I believe that the weak version is

[24] The same point is also persuasively pressed by Ducasse (1968a: 8; 1968b: 22) and put nicely by Dimock, who says that an innocent person can be "victimized" by the penal system but not "punished" by it (1997: 42).

[25] I am grateful to Jim Nickel for bringing this to my attention.

sufficient to generate the problem of punishment. It is difficult enough to justify the claim that it is morally permissible for the state to act with the intention of harming someone because he is believed to have broken the law. It doesn't matter whether or not the person believed to have broken the law did, in fact, do so.

1.1.5.2 The Consequentialist Objection

A second objection to the retributive requirement, so understood, maintains that it violates the neutrality condition that any reasonable definition of punishment must satisfy. The objection runs as follows. Some attempted justifications for the practice of punishment are essentially "backward-looking" in nature. The considerations they appeal to as a justification for punishment lie in facts about what was true prior to the time that the punishment is imposed. On one retributivist account, for example, punishing an offender now is justified because the offender deserves to be punished, and the offender deserves to be punished because, in the past, the offender committed a wrongful act. So, a retributivist of this sort would have no objection to the stipulation that to count as a punishment, the act must be done because the recipient of the punishment acted in a certain way in the past. But other attempted justifications for punishment are essentially "forward-looking" in nature. On a typical consequentialist account of punishment, for example, punishing an offender now is justified because it will deter the offender and others like her from committing such infractions in the future. A defender of such a theory might therefore object to the claim that to count as a punishment, the act must be done because of the offender's past behavior. Rather, on this account, it must be done because of future benefits that will accrue from doing it. And so, according to this consequentialist objection, the retributivist requirement makes the definition of punishment beg the question against consequentialist theories. This means that any definition that incorporates the requirement will fail the neutrality test and will therefore be unacceptable.[26]

This consequentialist objection to the retributive requirement fails because it neglects the distinction between the nature of punishment and the justification of punishment.[27] Consider again the concept of "repayment." Essential to its nature is that repayment involves paying someone back for a previously bestowed benefit. If I give you ten dollars and you have never provided me with any good or service, then my giving

[26] This objection is sometimes attributed to Quinton (1954: 59) [see, e.g., Lacey (1988: 7)], but it is worth noting that while Quinton does claim that the retributivist thesis is "an account of the meaning of the word 'punishment'" on this basis, it is not at all clear that he takes this to be an objection to the definition itself.

[27] Those who note that these are two different questions include Hart (1968: 263), Primorac [sic] (1981: 207–8), Primoratz (1989b: 188), and Adler (2000: 1414).

you the ten dollars cannot count as an act of repaying you. This is a claim about the nature of repayment. This claim, however, is completely neutral on the question of what, if anything, justifies the claim that repayment is a morally commendable way to behave. In particular, it is neutral about whether the practice is justified on backward-looking grounds (e.g., "because he is owed the money") or on forward-looking grounds (e.g., "because repaying people generates trust, which creates future benefits to society"). And the same is true of punishment. Legal punishment involves drawing a line on the basis of essentially backward-looking features – the line between those who have (or are believed to have) committed a legal offense in the past and those who have (or are believed to have) not – and then treating that line as morally relevant in determining how such people may permissibly be treated in the present. The claim that this is an essential feature of legal punishment, however, is neutral on the question of what, if anything, renders legal punishment morally permissible. And it is neutral on this further question precisely because it is neutral on the question of what, if anything, makes the line that it draws morally relevant. It is consistent with the retributivist view that the line is morally relevant because it separates those who deserve to suffer from those who do not, for example, but it is equally consistent with the consequentialist view that the line is morally relevant because of the good consequences that follow from drawing it there. Including the retributive requirement in our definition of punishment is therefore dictated by considerations of accuracy and illumination and is not overturned by considerations of neutrality.

1.1.6 Reprobative Retributive Intentional Harm

The result of the discussion to this point is that punishment involves acting with the intention of harming someone because she has (or is at least believed to have) committed an offense. These requirements are needed to distinguish accurately between cases of punishment and various cases that do not involve punishment. They do so in a manner that helps to illuminate what is distinctive about punishment. And they do so in a manner that is neutral on the question of whether or not punishment is justified and, if it is, on what grounds it is justified. There are two reasons, however, to hold that we are still short of an adequate definition. The first is that there are other differences between some of the pairs of cases previously discussed that these requirements fail to illuminate. The second is that there are other pairs of cases, not yet discussed, that these requirements fail to distinguish accurately. A definition based solely upon the requirements of harm, intention, and retribution, therefore, is still insufficiently illuminating and accurate.

To begin with the first consideration, let us return to the case of paying a fee when you voluntarily marry to one person versus paying a fine when

you voluntarily marry more than one person at the same time. A fine is clearly a punishment, while a fee is clearly not. Part of the difference between a fine and a fee is illuminated by attending to the distinction between intending harm and merely foreseeably bringing it about, and part of it is illuminated by appealing to the fact that a fee is charged for doing a legal act, while a fine is charged for doing an illegal one. But focusing exclusively on these two differences threatens to obscure a further relevant difference between the cases. When the state charges you a fee to process your wedding license, it is in no way expressing disapproval of your decision to get married. But when the state imposes a fine on you for violating antipolygamy laws, part of what it is doing is expressing its disapproval of your behavior. And the same is true of the other cases that in part helped to motivate the intentional harm requirement. When we imprison someone because he committed an offense, for example, we are in part admonishing him for his behavior. When we quarantine someone because she has contracted a highly contagious disease, on the other hand, we impose a similar kind of harm on her, but in no way do we mean to express disapproval of her. These considerations demonstrate that a fully illuminating definition of punishment must include a further requirement: to count as a punishment for an offense, the act must express official disapproval of the offender. As Duff and Garland put it, punishment "involves an essential element of condemnation" (1994: 13).[28] Call this the "reprobative requirement."

The second reason for accepting the reprobative requirement arises from consideration of a further case. If a definition of punishment limited to considerations of harm, intention, and retribution is insufficiently precise, then there will be cases in which an offender is harmed intentionally and because he committed a legally prohibited act but that are still not cases of punishment. I believe that there can be such cases and that, even though they are more contrived than those discussed to this point, they provide further valuable support for the reprobative requirement.

So, consider first the fact that there seem to be contexts in which the deliberate infliction of pain is meant to convey approval rather than disapproval. There is, for example, the practice of hazing within fraternities, as well as painful initiation rites within gangs and other clubs and organizations. As part of such a practice, the authorized leader of a group might deliberately inflict painful treatment on a new member, say by branding him with a hot iron, and the painfulness of the treatment might well be intentional and not merely foreseen. That is, if the treatment

[28] Others who emphasize this reprobative element of punishment include Hart (1958: 14–15), Charvet (1966: 577), Hawkins (1969: 124–5), Feinberg (1970: 75, 86), Husak (1990: 84ff.), Nino (1991: 260), Von Hirsch (1993: 9), Hill (1997: 196; 1998: 343), and Miller (2003: 33–4).

didn't cause pain to the recipient, then it would not fulfill its function as a rite of initiation (the iron, for example, would have to be made hotter if it didn't hurt the first time). But, at least in the sort of case I am concerned with here, the message conveyed by the harsh treatment would be one of approval rather than disapproval. This is how it is understood by the person who inflicts the pain, by the person on whom the pain is inflicted, and by the other members of the relevant group. Now consider the case of a gang that requires potential new members to break the law – say, by stealing a car – in order to be eligible for initiation. In this case, the leader of the gang who brands the new member with a hot iron after the new member steals a car deliberately inflicts painful treatment on the new member precisely because he broke the law. Clearly, this is not a case of the new member being punished for his offense. But, just as clearly, this assessment cannot be justified without adding the reprobative requirement to the definition of punishment.

The need for the reprobative requirement is perhaps less immediately apparent than the need for some of the other elements of our definition. But once the need becomes apparent, it is difficult to imagine a convincing objection to it. One might, at first, be tempted to object that the requirement begs the question in favor of one particular justification of punishment: that punishment is justified as a means of expressing or communicating society's disapproval of the offender's act. But, like the consequentialist objection to the retributive requirement considered in Section 1.1.5.2, this objection confuses the question of what punishment is with the question of what makes punishment permissible. The reprobative requirement maintains that part of what makes an act a punishment is that it expresses official disapproval of the offender's behavior. But this requirement is entirely neutral on the question of whether or not this feature of punishment, or any other feature of it, renders it morally permissible.[29]

1.1.7 Authorized Reprobative Retributive Intentional Harm

The discussion to this point has focused on how punishment affects the person punished and on what reasons are given for it. The question of who is doing the punishing has been passed over. Does this question matter? If we were looking for a definition of punishment in general, this question would potentially become somewhat complicated. Some philosophers have argued that if acts of the sort that I have described are carried out by private citizens, then they cannot qualify as punishment for that very reason, but only as something like vigilante justice. Primoratz, for example, has argued that "by definition, punishment is determined and executed by those *authorized* to do so" [(1989: 84); see also Benn)].

[29] This point is clearly made by Davis (1991a: 313), though in a slightly different context.

Others, however, have argued that such acts should still be construed as punishment, though perhaps as unauthorized and impermissible punishment. Still others have puzzled over the question of whether or not a person can punish himself or whether a wrongful act can, in effect, be its own punishment.[30] For our purposes, however, we can sidestep these particular problems. This book is concerned with the permissibility of legal punishment, and where I have used the term "punishment," it has been as shorthand for "legal punishment." And whatever we might think about the coherence or existence of other forms of punishment, it is clear that a punishment cannot be a legal punishment, in particular, unless it is carried out by an authorized agent of the state acting in his or her official capacity.[31] Call this the "authorization requirement."

The case for the authorization requirement, understood as a necessary condition for legal punishment, is overwhelming. In the absence of such authorization, there is no reason to think of an act as a legal punishment, even if there is reason to think of it as a punishment of some sort. The only objection to the requirement that I am aware of maintains that it begs the question in favor of the moral permissibility of punishment and thus violates the neutrality condition established as one of the conditions for an acceptable definition. But this objection rests on the assumption that if an act is legally authorized, then it is morally permissible. And this assumption is plainly false. The authorization requirement is therefore consistent with definitional neutrality.

It is possible, of course, that still further restrictions are needed. However, I am not aware of any considerations that would justify them.[32] At this point, then, our definitional work would seem to be complete. The results can be put more formally in terms of a stronger and a weaker set of conditions. On the stronger version, *P*'s act *a* is a legal punishment of *Q* for offense *o* if and only if

(1) *P* is a legally authorized official acting in his or her official capacity and
(2) *P* does *a* because *P correctly* believes that *Q* has committed *o* and
(3) *P* does *a* with the intent of harming *Q* and
(4) *P*'s doing *a does in fact* harm *Q* and
(5) *P*'s doing of *a* expresses official disapproval of *Q* for having committed *o*.

[30] See, e.g., the conflicting positions taken by Winch (1972), Paton (1979), and Reiff (2005: 27–8) on whether the agent's subsequent regret can itself constitute a punishment.

[31] For a useful examination of conceptions of punishment in a variety of nonlegal contexts, see McCloskey (1954).

[32] It might be argued, for example, that a person cannot be punished by an act unless she knows that it is intended as a punishment, but it could equally be maintained that in such cases the person is in fact punished but does not realize it.

On the weaker version, *P*'s act *a* is a legal punishment of *Q* for offense *o* if and only if

(1) *P* is a legally authorized official acting in his or her official capacity and
(2) *P* does *a* because *P* believes (perhaps mistakenly) that *Q* has committed *o* and
(3) *P* does *a* with the intent of harming *Q* (even if *P* fails to actually harm *Q*) and
(4) *P*'s doing of *a* expresses official disapproval of *Q* for having committed *o*.

I have argued for the cogency of the stronger version of the definition. If the recipient of the treatment in question is not guilty of an offense or is not harmed by the treatment, then I have maintained that he is not punished by it. For the purposes of this book, however, the differences between the weaker and stronger versions need not detain us, as long as the reader accepts one version or the other. As long as one agrees that attempting to punish people involves intending to select those who break the law and to harm them for doing so, the problem of punishment will arise with sufficient force. When, in the final analysis, we agree that legal punishment is authorized intentional reprobative retributive harm, therefore, I will mean that we at least accept the weaker version of the definition we have finally arrived at.

Allowing that punishment may be construed in terms of treatment aimed at those the state *believes* to have broken the law, however, raises one final problem that must briefly be addressed. For without two further restrictions, such a definition will include two problem cases that many defenders of punishment will not wish to defend. Since we want an account of legal punishment that will not beg the question against its permissibility, we must restrict the weaker version of the definition still further to prevent this from happening. First, there can be cases in which even though *P* believes that *Q* committed *o*, *P*'s belief is not reasonable. Suppose, for example, that a judge concludes that a defendant is guilty simply because the defendant is black. In this case, the judge might then send the defendant to prison because he believes that the defendant broke the law, and the weak version of the definition as it currently stands would include this as a legal punishment. But since it would not be fair to insist that a defense of legal punishment include a defense of acts such as that committed by the judge in this case, it would not be fair to include such acts within the scope of our final definition. Second, there can be cases in which even though *P*'s belief in *Q*'s guilt is reasonable, the evidence used in arriving at the belief is obtained at in morally objectionable ways. Suppose, for example, that a defendant really did commit a particular offense, but that the state proved its case by beating witnesses,

breaking into people's houses and offices, depriving the defendant of the right to speak to an attorney, and so on. Since it would again be unfair to insist that a satisfactory defense of legal punishment include a defense of punishing a defendant in such circumstances, it would again be unfair to include such acts within the scope of our definition. When the definition of punishment as authorized intentional reprobative retributive harm is understood in the weaker sense, therefore, its second clause must be amended to refer only to cases in which *P* justly and reasonably believes that *Q* committed *o*.

1.1.8 Punishment versus Restitution

I have argued that legal punishment should be defined as authorized intentional reprobative retributive harm. In arguing for this definition, I have claimed that it does the best job of accurately distinguishing between cases of punishment and other cases, and that it does so in an illuminating and neutral manner. In developing this argument, I have used examples of practices other than punishment, including curfews, quarantines, pretrial detention, gratuitous infliction of harm, and so forth. I have, however, deliberately refrained from appealing to one further case: compulsory victim restitution. I have done this precisely so that the definition of punishment arrived at will not have been influenced ahead of time by preconceptions about the relationship between punishment and restitution. Having arrived at a satisfactory definition of punishment, however, we must now ask: given this definition, what should we say about the practice of compelling an offender to make restitution?

Let us begin with an example. Larry vandalizes Moe's car by painting obscene words on it and breaking the windshield. Larry is caught and found guilty. The judge orders Larry to compensate Moe for the harm he has caused. She forces Larry to remove the spray paint from Moe's car, to pay the costs of replacing the windshield, to pay the costs involved in Moe's renting a car while his is in the shop, and to compensate him for the inconvenience and emotional distress Larry caused him. This seems clearly to be a case of compulsory victim restitution. Is it also a form of punishment?

Clearly, the act satisfies the authorization requirement. The judge is a legally authorized official acting in her official capacity. Clearly, it satisfies the retributive requirement. She orders Larry to do these things because Larry vandalized Moe's car. It seems equally clear that the judge's act satisfies the reprobative requirement. The act at the very least, expresses the view that Moe is entitled to have his car returned to its original condition, which means that Larry was not entitled to damage it, which in turn means that Larry did something that he was not entitled to do. And finally, it seems clear that the judge's act harms Larry. Larry is made

worse off by having to give Moe some of his money and spending some of his time cleaning up Moe's car. In all of these respects, the judge's act of forcing Larry to make restitution to Moe satisfies the definition of punishment we have arrived at. Indeed, for some theorists, these similarities between punishment and restitution seem to be sufficient to conclude that restitution is a form of punishment.[33]

But this conclusion is premature. For one further question remains: in imposing this burden on Larry, does the judge act with the intention of harming Larry? As we saw in Section 1.1.4, distinguishing between harm and intentional harm is necessary in order to account fully for the difference between, for example, fees and fines. If the answer to this question is yes, then restitution is more like a fine than a fee and is a punishment. If the answer is no, then restitution is more like a fee than a fine and is not a punishment.[34]

The answer to the question is this: there is no one answer that covers every case in which a judge compels an offender to make restitution to his or her victim. In some cases, the judge's intent may be to impose a cost on the offender. She may say, for example, that she wants Larry to suffer the drudgery involved in cleaning the paint off Moe's car so that Larry will come to see how wrong his act was. In this case, part of the judge's motivation for imposing the cost on Larry is, indeed, punitive. Following Barnett, we can refer to cases of this sort as cases of "punitive restitution" (1977: 219–20). But this need not be true of all cases. In some cases, the judge's reasoning may be that Larry must pay the money because he owes it to Moe. In this case, the judge foresees that paying Moe will impose a

[33] Martin, for example, characterizes "compensation, to be paid to victims of crime" as one of the three main "modes of punishment": "[w]hen a violator is *forced* to pay compensation – as a feature of the verdict against him – such compensation counts as punishment" [(1987: 73, 74); for a related argument, see Martin (1990)]. Similarly, Abel and Marsh argue that compulsory victim restitution is a punishment "because it involves unpleasant consequences for an individual who has interfered with our pursuit or realization of individual and social ends" (1984: 18–19). Similar views are endorsed by Baylis (1968a: 45–6), Dagger (1991: 36–7), Hoekema (1991: 336–40) and, at least implicitly, Loewy (2000: 15–16).

[34] Abel and Marsh maintain that restitution is a punishment in part because they deny the intentional harm requirement. They point out that forbidding someone to work in a certain profession can be a form of punishment and that the state can establish licensing requirements "which have the effect (though not the purpose) of denying particular members of affected professions the right to practice their calling." From this, they conclude that "the state may punish, though its intent is to regulate," (1984: 39) and that the intentional harm requirement should therefore be rejected. But it is independently objectionable to hold that licensing requirements are a form of punishment. If they are, after all, then they clearly amount to punishing the innocent and yet are widely viewed as morally permissible. It is therefore much more plausible to conclude that licensing requirements are not a form of punishment than to assume that they are and conclude that punishment need not involve the intent to cause harm to the one punished.

cost on Larry in the same way that she might foresee that enforcing the terms of a contract will impose costs on one of the parties to the contract, but this fact plays no role in her decision. Following Barnett, we can refer to cases of this sort as cases of "pure restitution" (1977: 220). An argument against the moral permissibility of punishment, therefore, will also provide an argument against punitive restitution. But it will provide no reason to reject pure restitution.[35] And this fact will become important toward the end of our investigation.

1.2 WHAT THE PROBLEM OF PUNISHMENT IS

1.2.1 The Problem

Let us now assume that we have a sufficiently clear understanding of what legal punishment is. Given this understanding, what is the problem? The answer is simple. First, punishment involves drawing a line between two different sets of people and treating the members of one group very differently from the members of the other. Surely, in general, it is wrong to treat two groups of people so differently unless there is a morally relevant difference between them. So, part of the problem is explaining why the difference between those that punishment targets and those that it does not is morally relevant. Second, punishment involves not merely treating the members of one group differently from the members of the other, but harming the members of one group and not those of the other. The fact that an act will harm someone is clearly morally relevant, so part of the problem is explaining why the difference between offender and nonoffender is important enough to justify acts that will harm the offender. Finally, and perhaps most importantly, punishment involves not merely acts that predictably harm offenders, but acts that are carried out precisely in order to harm them. Since it is considerably more difficult to justify intentionally harming someone than it is to justify merely foreseeably harming her, the problem of punishment is even greater than it might at first seem: we must explain not only why the line between offenders and nonoffenders is morally relevant at all but, in particular, how it can be important enough to justify not merely harming those on one side of the line but intentionally harming them. Punishment, in

[35] Barnett distinguishes between pure and punitive restitution in, e.g., (1998: 204). If it seems odd that the permissibility of the state's act could turn on the intention of the official who carries out the act on the state's behalf, it should be noted that this would seem to be true regardless of what one thinks of the relationship between punishment and restitution. It could, for example, be permissible for an official to approve the construction of a new airport while foreseeing that it will cause headaches to many people who live near the construction site but be impermissible to approve construction of the airport in order to cause the headaches.

short, involves the state's treating some of its citizens in ways that it would clearly be wrong to treat others. The problem is to explain how this can be morally permissible.

There have, of course, been many attempts to do this, and I will attempt to provide a critical response to them in the chapters that follow. Before doing so, however, I want to attend to two final preliminary considerations. First, I want to respond to a few reasons that have been given for thinking that there is no need for the defender of punishment to shoulder this burden. These reasons attempt to support the view that the problem of punishment is a pseudoproblem. Second, I want to clarify what, precisely, the burden on the defender of punishment amounts to. This means explaining what a response to the problem of punishment would have to accomplish in order to count as a successful solution to the problem.

1.2.2 The "Principle of Separability" Argument

One reason that might be given for denying that punishment requires moral justification rests on what Martin calls "a methodological principle of separability." On this principle, politics and morals "are logically distinct. Politics is not essentially, or logically, a special case of morals; and political theory is not an applied form of ethics" (1970: 254). Since legal punishment is an essentially political institution, it follows that on this account "punishment does require a political justification. But this is the only kind it *needs*" (255). To justify punishment politically is to show "its 'necessity' within a particular system of *political* institutions and principles," and "[o]nce the job of political justification is done," Martin urges, "it is difficult to see any role for independent moral judgment" (254, 255).[36] The claim that legal punishment raises a moral problem, on this account, rests on a category mistake: it involves the attempt to subject to moral scrutiny a practice that, by its very nature, is not to be subject to such scrutiny.

The principle of separability argument must be rejected for two reasons. First, it rests on a confusion between two different senses of "justification." In one sense, to justify a practice is to establish that it is morally permissible. In a second sense, to justify a practice is to establish that it serves a good purpose. A practice can be justifiable in the first sense but not the second, and can be justifiable in the second sense but not the first. Making a practice of always putting one's left shoe on before one's right, for example, is surely permissible, but it serves no good purpose. Making a practice of performing painful medical experiments on unwilling human subjects might serve a good purpose but is morally impermissible nonetheless.

[36] For a more modest version of this argument, see Philips (1986); for a similar position, see Blumoff (2001: 166–8).

The distinction between these two senses of justification is important for the following reason. Martin's defense of the claim that punishment needs a political justification but not a moral one is a defense of a claim about justification in the second sense. Since legal punishment is a political institution, that is, the argument maintains that one can show that it has a purpose only by showing what its purpose is from within the political system as a whole. Indeed, Martin at one point writes that "[t]o justify punishment is to display its *rationale*" (254), and this is clearly to speak of justification in the second sense. But the problem of punishment is concerned with justification in the first sense. The fact that punishment involves the state's deliberately inflicting harm on some of its citizens, that is, does not raise the question "does this practice serve a good purpose?" but rather the more fundamental question "is this practice morally permissible in the first place?" Once this distinction is recognized, it should become clear that the principle of separability argument provides no good reason to doubt that an accurate understanding of the nature of legal punishment gives rise to a genuine problem about its moral justification.

The second reason for rejecting the principle of separability argument takes the form of a reductio ad absurdum. The argument can succeed only if it is true that so long as a political institution is politically justified, its moral permissibility does not require justification. But if this were true, then all sorts of clearly objectionable practices could not be criticized on moral grounds. The extermination of ethnic minorities in concentration camps, for example, could clearly be "justified" in the sense that it could have a "rationale" from within a given political system, even a system in which political leaders are democratically elected. If the principle of separability argument is correct, then, we cannot criticize such a practice as morally impermissible. But this is plainly unacceptable. And so too, therefore, is the principle of separability argument.

1.2.3 The Logical Entailment Argument

A second argument against the claim that punishment generates a genuine moral problem maintains that there is no need to provide a moral justification for the state's right to punish those who violate the law because this right follows logically from the concept of a legal requirement. This position is most fully developed in Herbert Fingarette's article "Punishment and Suffering" (1978).[37]

[37] Although a closely related argument is developed throughout Reiff (2005). See also Murphy, who has written that "It seems plausible to maintain that the appropriateness of punishment follows *logically* from the recognition of certain human acts as violations of the criminal law, which is just that branch of law mandating punishment" (1990a: 1, emphasis added). Similarly, Morris has written that "*Logically* connected with the concept of wrongdoing is the concept of a painful response that another is entitled to inflict because of

Fingarette's argument runs as follows: "If, when I will contrary to law, the law can or will do nothing, then ... so far as the law is concerned my will is unqualifiedly *un*constrained in this regard." But, Fingarette continues, "the very point of a legal requirement upon me is that I *am* constrained, *un*free, *required* to obey." To say that there is a legal requirement that I behave in a certain way, therefore, is to say that if I don't behave in this way, I will "suffer some constraint upon my will. Anything less than this makes the notion of 'requirement,' as distinguished from 'request,' unintelligible" (1977: 509).[38] And this, according to Fingarette, is just to say that punishment follows, as a simple matter of logical entailment, from the idea of a legal requirement: "the necessity for punishment is independent of moral justification," that is, because "the necessity is internal to law" itself.

There are two reasons to reject the logical entailment argument. The first is that the entailment claim itself is false. Requirements in general, and legal requirements in particular, are perfectly intelligible in the absence of punishment. Fingarette's argument for the claim that they are unintelligible moves from the claim that if the state does nothing in response to my violating the law, then it is not really requiring me to behave in a certain way to the claim that if the state does not punish me in response to my violating the law, then it is not really requiring me to behave in a certain way. But there is a large gap between not punishing and not doing anything. And there are many nonpunitive ways to fill this gap. The state, for example, could publicly denounce offenders for breaking the law.[39] It could banish them.[40] It could ostracize them.[41] It could compel them to make restitution to their victims.[42] It could require them to seek psychological treatment or education.[43] It could

the wrongful conduct" (1981: 47, emphasis added). Primoratz (1989a: 152) also seems to endorse this view, as do Andenaes (1970: 131), Newman (1978: 192; 1983: 100), and Nathanson (1985: 191) [see also Gaffney (1994: 13–15) for a very similar argument].

[38] Similarly: it "is not contingent ... that those who disobey requirements must be made to suffer" (1977: 513); legal requirements "can only be intelligible in the institutional framework of power if there is a policy of retributive punishment" (514).

[39] Ten (1987: 40–1), for example, makes this suggestion.

[40] Banishment, of course, can also be used as a punishment. If a person is banished to make him suffer, then he is punished by being banished. But if he is banished simply to prevent others from being harmed by him, then, as in the case of a quarantine, this is not a form of punishment.

[41] Bedau (1978: 618), for example, notes that if you break the rules of chess in playing against me, it does not follow that I have a right to punish you, though I might well be entitled to refuse to continue to play with you.

[42] Wright points out that historically, some societies have used a restorative rather than a punitive response to violations of legal requirements (1996: 11, 19). See also Barnett (1977).

[43] Johnson (1985: 159) stresses this option in his response to Fingarette. See also Stephenson (1990: 229).

detain them as a purely preventive measure. And, perhaps most importantly, it could actively seek to prevent them from breaking the law in the first place. All of these are ways in which the state could respond to violations of the law without punishing people that would be logically consistent with its maintaining that obeying the law is required of its citizens and not merely recommended. This is not to say that these are the best ways for the state to respond to legal violations or even that they are particularly desirable ways. It is simply to say that they are logically possible ways. But saying this is enough to say that the logical entailment argument must be rejected.

This first argument against the entailment claim can usefully be supplemented by considering its application in nonlegal contexts. Consider, for example, a museum that does not charge an admission fee but does publicize a suggested donation of five dollars per visitor. It also notifies visitors of a policy by which people are admitted only if they are wearing a shirt and shoes. The first policy amounts to a request, the second to a requirement. Now the proponent of the logical entailment argument is clearly correct to maintain that the distinction between request and requirement will be unintelligible if the museum treats those who try to get in without paying five dollars in the same way that it treats those who try to get in without a shirt and shoes. If you enter the museum without making the requested donation, you will still be entitled to do everything that those who do make a donation are entitled to do. This is because the donation is merely a request. If the museum treated those who entered without shirts or shoes in precisely the same way, then it would seem to be correct to say that the rule about shirts and shoes was not a requirement in the first place. But surely this does not mean that the museum must *punish* people who try to get in without a shirt or shoes. It simply means that it should prevent such people from entering the museum and remove them, perhaps by force if necessary, if they get in undetected. So long as the museum behaved in this manner, no one would doubt that the rule about shirts and shoes was a requirement, while the rule about giving five dollars was not. And since the museum would not be punishing people for violating the rule, this example again demonstrates that a requirement does not logically entail punishment.[44]

[44] A further consideration against the entailment claim lies in the domain of etiquette. It is a rule of etiquette, for example, that one must respond promptly to an invitation. There is nothing amiss about insisting that this rule is a genuine requirement of etiquette, and not a mere suggestion, while at the same time recognizing that etiquette does not require (indeed, does not even permit) one to punish someone for breaking the rule. There are, of course, important differences between rules of law and rules of etiquette, but what they have in common is sufficient to establish the relevant point: a rule can be a genuine requirement within its domain without entailing that within its domain violators of the rule ought to be punished.

In addition, there is at least one important legal counterexample to the entailment claim. In many countries, including the United States, it is or has been illegal to commit suicide. But if a person does kill himself, the state does not punish him for violating this legal requirement. Indeed, not only does the state *not* punish people who commit suicide, the state *cannot* punish such people. And, even more importantly, people who contemplate committing suicide are fully aware of this ahead of time. There is nothing incoherent or unintelligible about the state's forbidding suicide and at the same time recognizing that it cannot punish people for violating the prohibition. And so, once again, there is no reason to think that the existence of a legal requirement logically entails the existence of a practice of punishing people who violate it.

But let us now suppose that the entailment claim can somehow be sustained. Even if this is so, the argument from this claim to the conclusion that the problem of punishment is a pseudoproblem must still be rejected. For the argument will now simply beg the question: how can it be morally permissible for the state to issue legal requirements in the first place?[45] If issuing a legal requirement means issuing a specific kind of rule, one that commits one to intentionally harm those who violate it, then what makes it permissible to issue this kind of rule in the first place? Considerations of logic alone cannot answer this question, and so considerations of logic, in the end, can do nothing to cast doubt on the claim that the state's practice of punishing people for breaking the law requires moral justification.[46]

[45] Duff (1986: 197) notes this problem with the logical entailment argument. Fingarette at one point seems to concede this point (1977: 513), but conceding it would deprive the argument of any real force.

[46] A related entailment argument in defense of punishment has been made by a few other writers. On this account, it is not the existence of a legal requirement but rather the existence of personal responsibility that logically entails the permissibility of punishment. As Oldenquist puts the claim, for example, "Personal accountability *makes no sense* unless it means that transgressors deserve punishment – that is, that they are owed retribution" (1986: 76, emphasis added). See also Oldenquist (1988), in which he maintains that it is "true by definition" that holding people personally accountable for harms they cause "is to consider them deserving of blame and punishment" (466, 465). The same position is also defended by Barton (1999: 93–7). But this entailment argument fails for precisely the same two reasons that the original one does: the entailment claim itself is false (we can hold people responsible for the harms they wrongfully cause by morally criticizing them, banishing them, forcing them to make restitution to their victims, etc., none of which involve the intentional infliction of harm), and if it is simply stipulated to be true, then the argument begs the question: why is it permissible to hold people responsible for their behavior in this very narrow sense if doing so simply means treating them in the ways that punishment involves treating people?

1.2.4 Testing Solutions to the Problem

I have argued that the problem of punishment is, indeed, a problem and that it does, indeed, require a solution. Let us now suppose that I am correct. If the moral permissibility of legal punishment does pose a problem, then a successful solution will have to do two things. It will have to provide a moral principle or set of moral principles to serve as a foundation for justifying punishment, and it will have to demonstrate that the foundational principle or set of principles does, in fact, support the claim that punishing people for breaking the law is morally permissible. There are therefore two distinct tests that any proposed solution to the problem of punishment must pass. First, it must pass what I will call the "foundational test." This involves examining the solution's foundational principle or set of principles to determine whether or not it is acceptable on its own. Second, it must pass what I will call the "entailment test." This involves examining the route from the foundational principle or set of principles to the conclusion that the practice of punishment is morally permissible to see whether the inference from the one to the other is itself acceptable. A solution to the problem of punishment must pass both of these tests in order to be successful.

1.2.4.1 The Foundational Test

The principal means of employing the foundational test consists of probing a solution's foundational principle or set of principles to see whether it has any implications that would render it unacceptable. And although a particular implication could prove to be unacceptable because it was literally incoherent (a set of principles might be turn out to be jointly inconsistent, for example), in practice, this means of testing a solution almost always involves evaluating its substantive implications according to our considered moral intuitions. If the implications of a solution's foundation are sufficiently counterintuitive, this will lead most people to reject the foundation itself, and will thus count, at least for them, as a reason to declare that the solution has failed the foundational test. And this, in turn, will render the solution unsuccessful. If a particular moral principle is unacceptable in itself, after all, then the fact that it would justify punishment if it were acceptable will do nothing to justify punishment.

Whether or not a given implication of a particular principle is counterintuitive, of course, can vary from person to person, as can the question of how strongly counterintuitive it is and how much weight should be put on an implication's counterintuitiveness in the first place. It is therefore unlikely that the foundational test can be used decisively to reject any particular solution to the problem of punishment in a manner that will prove sufficiently compelling to everyone. There are, nonetheless, some

relatively basic intuitions that most people seem to put a good deal of weight on and that can thus be used as the basis for what I will refer to in this book as a "reasonable application" of the foundational test. Virtually everyone who has attempted to justify punishment, for example, firmly believes that punishment should be at least roughly proportionate to the severity of the offense. A foundational principle whose implications are too strongly at odds with this proportionality requirement (permitting, say, the painful execution of people for relatively minor traffic offenses or permitting no more than a modest fine for child molesters) may thus reasonably be deemed to have failed the foundational test even if some people would be willing to accept such implications rather than reject the theories that give rise to them. For the purposes of this book, therefore, I will take it that a proposed solution to the problem of punishment fails a reasonable application of the foundational test if it has a substantive implication that a large majority of people, including a large majority of those who believe in punishment, firmly reject.

1.2.4.2 The Entailment Test

While the foundational test involves an appeal to intuitions that can vary from person to person, however, the entailment test does not. The entailment test simply asks whether or not any given foundational principle or set of principles would, if it were true, establish that punishment was morally permissible. The entailment test can thus be used in a manner that should prove equally effective to everyone, regardless of particular intuitions about particular cases. As we saw in Section 1.1.4, for example, punishment involves not merely harm but intentional harm. A foundational principle that could justify an act that would harm an offender but that could not justify acting with the intent of harming her, therefore, would fail the entailment test.[47] And, as we saw in Section 1.1.5, punishment treats the line between those who have broken the law and those who have not as morally relevant in determining who may permissibly be subject to such harm. If a moral principle cannot justify treating the line between offenders and nonoffenders as relevant in the way that punishment treats that line, then the principle will again fail the entailment test. If a moral theory turned out to entail that all human beings should be intentionally harmed, for example, then it would, as a consequence, entail that all offenders should be intentionally harmed, but this would not be the same thing as justifying punishment. The fact that a given foundational moral principle fails the entailment test does not in itself provide a reason to abandon the principle itself. If the truth of act-utilitarianism does not support the truth of the claim that punishment is

[47] For examples of solutions to the problem of punishment that are ultimately unsuccessful at least in part for this reason, see Sections 3.2.9, 3.3.7, 3.4.1.3, 4.1.5, and 4.4.5.

permissible, for example, this would not be a reason to reject act-utilitarianism. But even if a moral principle that fails the entailment test is not rejected as a moral principle, if the principle cannot ground a justification for intentionally harming those who break the law, then whatever else it may justify, the principle cannot justify legal punishment and thus cannot serve as a basis for a successful solution to the problem of punishment.

The goals of this preliminary chapter have been essentially clarificatory and somewhat modest. If they have been met successfully, it should now be clear what legal punishment is, what the problem of punishment is, and what tests a given solution to the problem must pass in order to be considered successful. The goal of the remainder of this book is substantially more than clarificatory and substantially less than modest: to establish that none of the proposed solutions in the extensive literature on the problem pass both of these tests (Chapters 2–4) and to argue that, since this is so, punishment should be abolished (Chapter 5).

2

The Consequentialist Solution

2.0 OVERVIEW

The consequentialist solution to the problem of punishment maintains that punishing people for breaking the law is morally permissible because of its presumed good consequences. This solution can be defended in two importantly distinct ways. First, the consequentialist solution in particular can be defended in the context of a broader defense of consequentialism as a moral theory in general. On this account, any behavior is morally justified if and only if its expected consequences are better than those of any available alternative behavior and punishment is simply one particular instance of this general phenomenon. A second defense of the consequentialist solution seeks to defend punishment in particular without committing itself to this larger principle about morality in general. On this second approach, the defender of punishment maintains that some things are morally justified by their positive consequences, and that punishment is one of these things, without insisting that everything that is morally justified is justified by its positive consequences or that everything that has positive consequences is thereby morally justified. Thus, while a successful refutation of consequentialism in general would count as a successful refutation of the broad defense of the consequentialist solution, it would not count as a successful refutation of the narrow defense of punishment in particular. And while an objection to the consequentialist justification of punishment in particular might show that the consequentialist solution has implications that are unacceptable to proponents of the narrow defense, it might turn out that proponents of the broader defense would be willing to live with them. A satisfactory response to the consequentialist solution must therefore address both points of view, showing that the solution is unacceptable to consequentialists and nonconsequentialists alike. The goal of this chapter is to do precisely that.

The claim that punishment is permissible because of its positive consequences can be understood in many different ways. What makes a

particular consequence positive rather than negative can be understood in different ways (in terms of maximizing pleasure, preference satisfaction, equality, etc.), and what is rendered permissible by its consequences can also be understood in different ways (acts, rules, dispositions, etc.). I will begin, in Section 2.1, by considering the version of the consequentialist solution that seems to me to be the most straightforward in both of these respects: one on which better consequences are understood in terms of promoting human well-being and on which individual acts are justified by bringing about such consequences. This position represents the act-utilitarian version of the consequentialist solution to the problem of punishment associated with such writers as Beccaria and Bentham and, more recently, Smart, and it will be explained in Section 2.1.1. I will then present a series of objections to the act-utilitarian solution, each of which has the form of a reductio ad absurdum. Each objection maintains that the act-utilitarian solution has an implication that is sufficiently objectionable to warrant rejecting the solution itself. Some of these objections (Sections 2.1.4–2.1.6) rest entirely on the assumption that the implications in question are deeply counterintuitive. These objections maintain that the act-utilitarian solution must be rejected because it does not pass what I called a reasonable application of the foundational test that was explained and justified in Section 1.2.4.1. They should provide the typical nonconsequentialist, at least, with sufficient reason to reject the narrow defense of the act-utilitarian solution, but they may well prove impotent when exploited against a sufficiently determined proponent of the broad defense of the solution who may be willing to stick with consequentialism as a whole regardless, as it were, of the consequences. But while this is true of some of the objections that I will raise in Section 2.1, it is not true of all of them. In Sections 2.1.2 and 2.1.3, in particular, I will argue that at least some of the implications of the act-utilitarian solution are also unacceptable, at least in part, for reasons that have nothing to do with our moral intuitions. These reasons have to do only with the nature of punishment itself, and they will therefore establish that even if the act-utilitarian solution passes the foundational test, it still fails the entailment test that was explained and justified in Section 1.2.4.2 (although, for most people, the considerations in these sections will also establish that the solution fails the foundational test for still other reasons). Section 2.1 as a whole, then, will demonstrate that the act-utilitarian version of the consequentialist solution to the problem of punishment is unacceptable regardless of whether or not act-utilitarianism as a moral theory in general is true. In Section 2.2, I will consider whether a rule-utilitarian version of the solution might be able to overcome the objections that undermine the act-utilitarian version. I will argue that it cannot. In Section 2.3, I will briefly consider the appeal to still further variants of utilitarianism and will argue that there is no reason to believe that any of

them will fare any better. Finally, in Section 2.4, I will consider a variety of nonutilitarian versions of consequentialism, and will argue that none of them can provide a successful solution to the problem of punishment either. The chapter as a whole, therefore, will demonstrate that the problem of punishment cannot be solved by appealing to the good consequences that punishment is presumed to bring about.

2.1 THE ACT-UTILITARIAN VERSION

Utilitarianism is a form of consequentialism that identifies the good to be promoted with human happiness or well-being. Act-utilitarianism is a form of utilitarianism that identifies individual human actions as the subjects to be evaluated according to how effectively they promote (or can reasonably be expected to promote) this good. According to the act-utilitarian solution to the problem of punishment, punishing people for breaking the law is justified because it best promotes (or can reasonably be expected to best promote) human happiness or well-being. This solution to the problem of punishment initially seems to be quite attractive. It seems right, in general, to suppose that punishing people for breaking the law is useful. And it seems right, in addition, to suppose that punishment would lose much of its intuitive appeal if could be shown to serve no useful purpose. But while the act-utilitarian solution does possess a certain surface plausibility, it is also subject to a variety of objections that ultimately render it unsuccessful.

2.1.1 The Act-Utilitarian Solution

Let us begin with a typical case of punishment. Larry robs a liquor store and is sentenced to five years in prison. It seems reasonable to expect that incarcerating Larry will have several significant consequences. First, it will significantly reduce his ability to commit further offenses while he is in prison.[1] Second, imprisonment may reasonably be expected to reduce the likelihood that Larry will commit more offenses after he is released. Finally, after learning of Larry's experience, other potential lawbreakers may also be expected to be less likely to commit offenses in the future. In all of these ways, it is reasonable to suppose that imprisoning Larry for his offense will reduce the amount of lawbreaking in the future.[2] In addition,

[1] This feature of the consequentialist solution is typically overstated as the claim that imprisoning an offender will prevent him from committing offenses while he is in prison. Strictly speaking, it is still possible (and in some cases, perhaps, probable) that an inmate will commit more offenses while in prison, either against his fellow inmates or, say by telephone, against people outside the prison. In addition, while incarceration may have this effect, this is not so of other forms of punishment such as monetary fines and corporal punishment.

[2] To say that it is reasonable to suppose this is not, of course, to insist that it is true. For the purposes of this chapter, I will assume for the sake of argument that punishment typically

it seems reasonable to suppose that it is better, from the standpoint of human happiness, when the amount of lawbreaking is reduced, both because legal offenses cause unhappiness for the victims and because such offenses typically have further consequences that, in turn, cause further unhappiness for others.

Of course, it is also true that incarcerating Larry will have other consequences that may reasonably be viewed as bad from the standpoint of human happiness, chiefly the consequences for Larry himself. But even taking into account the costs to Larry and to those who may miss him while he is in prison, the taxpayers who will pay to feed and shelter him, and so forth, it may seem plausible to suppose that, on the whole, the benefits in terms of overall human happiness generated by imprisoning Larry outweigh the costs and produce a greater overall balance of happiness over unhappiness than would responding to his offense in other ways or not responding at all. In short, it seems reasonable to expect that imprisoning Larry for his offense will have, on balance, better consequences than will not punishing Larry from a utilitarian point of view. Finally, it seems plausible to say that imprisoning Larry, which would be justified by these beneficial consequences, is an act of legal punishment, and not merely an act that is in some respects similar to an act of punishment.

All the features of the act of putting Larry in prison that make the act satisfy the definition of punishment defended in Section 1.1, that is, are features that seem to contribute to the positive consequences of the act. The act must satisfy the intentional harm requirement to serve as a useful deterrent: if we did not intend to harm Larry by putting him in prison, we could not intend to deter others from breaking the law by imprisoning him. The act must satisfy the retributivist requirement because, if the harm to Larry were not done in response to his having committed an offense (if, for example, Larry was simply chosen at random to be subjected to this harsh treatment), then, again, harming Larry now wouldn't deter anyone from committing offenses in the future. The act must satisfy the reprobative requirement because an act that conveys an expression of social disapproval will have more deterrent force than one that does not. And the act must be carried out by a recognized legal authority because the power of the state will be more intimidating to prospective lawbreakers than the much weaker power of isolated private individuals. And so, this act-utilitarian version of the consequentialist solution maintains,

has a significant deterrent effect, but it is worth noting that this is an empirical claim that would require justification. And, as Braithwaite and Pettit put it, it seems "fairly uncontroversial" to say that the "massive" literature on deterrence "failed to produce . . . evidence that more police, more prisons, and more certain and severe punishment made a significant difference to the crime rate" (1990: 2). For an especially effective critique of the claim that punishing people who break the law generally produces more utility than other alternatives, see Golash (2005: 22–43).

punishing Larry for his offense is morally justified. And since the considerations in defense of this claim about Larry seem to hold about offenders in general, and not just about Larry in particular, it follows that punishing people for breaking the law in general is morally permissible. The act-utilitarian solution to the problem of punishment is thereby said to be vindicated.

2.1.2 The Punishing the Innocent Objection

The most common and powerful objection to the act-utilitarian solution is as follows. If punishing Larry for robbing a liquor store is morally justified because it produces more overall utility than would any available alternative, then in at least some circumstances, deliberately punishing an innocent person is also morally justified because it produces more overall utility than would any available alternative. But since this implication is plainly unacceptable, so, too, is the act-utilitarian solution to the problem of punishment. I will refer to this as the "punishing the innocent objection."[3]

Examples designed to establish the soundness of the punishing the innocent objection to consequentialist theories of punishment are ubiquitous in the literature on punishment.[4] One can imagine, for example, cases in which the state must deliberately punish an innocent person who is widely believed to be guilty in order to prevent a riot from occurring if he is acquitted. There can be cases in which the state frames an innocent person and punishes her for a particular offense in order to deter others from committing that offense. A state might determine that punishing both an offender and his children or other relatives would more effectively deter others from committing such offenses in the future than would merely punishing the offender himself. The authorities might frame and punish an innocent person for a particular offense to tempt the actual perpetrators to become less cautious and to be caught in the future, and so on. There seems to be no shortage of scenarios in which the state

[3] For purposes of this discussion, I will set aside the question of whether persons should be considered innocent of an offense when they have not yet committed the offense but it seems clear beyond a reasonable doubt that they will commit it in the future. New, in particular, has argued that "prepunishing" a person in such cases (particularly when it is believed that it will be impossible to punish him after the fact) "is not in a morally significant way punishing the innocent at all" [(1992: 37); however, see Smilansky (1994) and New's response to Smilansky in (1995), as well as Statman's defense and extension of New's proposal in (1997)]. Since this is a contested question, I will stipulate that in the cases to be considered here, the innocent people have not committed the offense that they are punished for and there is no reason to believe that they will do so in the future.

[4] See, e.g., Mabbott (1939: 39), Hawkins (1944: 14), Lewis (1949: 305), McCloskey (1957: 468–9; 1967: 93–102), Armstrong (1961: 152), Brainthwaite and Pettit (1990: 46), Gavison (1991: 352, 256), Cragg (1992: 48–51), and Golash (2005: 43–4).

could, on the whole, do more good by punishing an innocent person than by not punishing him.[5]

According to the proponent of the act-utilitarian solution, the right act for you to perform in these cases (as in every case) is the one with the highest overall expected utility. An act-utilitarian would, of course, take into account the interests of the innocent person. But she would also have to weight these interests equally with those of everyone else. And once this is done, it will become clear that, on the act-utilitarian account, it will be morally permissible to punish an innocent person for an offense he did not commit. Indeed, strictly speaking, the act-utilitarian will have to conclude not merely that doing so would be permissible, but that the state would have a positive duty to punish the innocent person, that it would be positively impermissible for the state to refrain from doing this. Even more strikingly, it will follow that the innocent person himself will have a moral duty to go along with the charade (to falsely confess to the offense, to waive his right to an appeal, etc.) in order to contribute maximally to the social good.[6] And the act-utilitarian solution will entail all of this not only in cases where intentionally punishing an innocent person would be necessary to avert a catastrophe, but in every instance in which punishing an innocent person would, on the whole, produce even a little more utility than not doing so. But surely this cannot be right. And so, the act-utilitarian solution to the problem of punishment cannot be right either.

2.1.2.1 The Definitional Response

The punishing the innocent objection is an objection by reductio ad absurdum. There are therefore two principal ways in which a defender of the act-utilitarian solution might attempt to respond to it: she could try to show that there are good act-utilitarian reasons to refrain from punishing the innocent person, so that the act-utilitarian solution does not have this implication, or she could try to bite the bullet and admit that her solution to the problem of punishment does have this implication but deny that this should be a sufficient reason to reject it. I will consider each of these responses in the following two subsections. But first, it is important to acknowledge that according to some philosophers, the act-utilitarian need not rely on either of these responses and the punishing the innocent objection can be overturned simply by attending more carefully to the definition of punishment. I will therefore begin my treatment of the

[5] In addition, as Smilansky has pointed out, generally more overall good could be produced at the expense of punishing more innocent people by relaxing the commonly accepted standards of evidence, the degree of certainty required for a conviction, and so on, that constrain our punishment of the guilty because we currently put so much weight on preventing punishment of the innocent (1990: 258–61).

[6] Primoratz (1989a: 44) presses this point effectively (see also Primorac [sic] 1978).

punishing the innocent objection by explaining why this third response is unacceptable.

As I argued in section 1.1.5, part of what makes an act a punishment, and not merely a gratuitous infliction of harm, is that punishment involves harming someone because he has done a legally prohibited act. And so, it would seem, if a person has not done a legally prohibited act, then he cannot, by definition, be punished. And if he cannot, by definition, be punished, then it cannot be an implication of the act-utilitarian solution (or of any solution) that he should be punished. Benn, for example, famously offered this claim as a retort to the punishing the innocent objection: "The short answer to the critics of utilitarian theories of punishment, is that they are theories of punishment, not of any sort of technique involving suffering" (1958: 332).[7] I will refer to this as the "definitional response."

One possible reply to the definitional response would be to insist that it rests on a mistaken definition of punishment. Ten, for example, argues that "[t]he element of truth in [the definitional response] is that there must be some wrongdoing or some offense for there to be punishment. But this is not to say that the person punished must be the offender. An innocent person can be punished for an offense committed by someone else" (1987: 16).[8] According to the account of punishment arrived at by the end of Section 1.1, this reply to the definitional response might prove successful. On that account, we are free to choose between a stronger and a weaker version of the definition, and the weaker version does not insist that a person actually be guilty of an offense to be punished for it; it merely requires that he be believed to be guilty. And so, on this account, it is perfectly possible for an innocent person to be punished and therefore perfectly possible for the act-utilitarian solution to entail that innocent people should be punished. As I noted in Section 1.1.5.1, however, I favor the stronger version of the definition, on which, if a person has not committed an offense, then he cannot, strictly speaking, be punished. Just as someone who did not do a good act is not being repaid if the state benefits him under the false pretense that he did a good act, a person who is harmed under the false pretense that he did a bad act is not being punished. On this point, at least, the definitional response seems to me to be correct. And so, while the disjunctive definition I have agreed to work

[7] This is what Hart refers to as the "definitional stop" [Hart (1959): 5–6]. This definitional response can also be attributed to Bentham: "If one examines Bentham's writings on punishment and the discussions surrounding them, it soon becomes obvious that punishing the innocent was not originally envisaged as a problem in his theory. Punishment follows the breaking of legal rules, and hence, punishment *implies* guilt in this sense" [Rosen (1997: 25, emphasis added)].

[8] The same argument against the necessary connection between punishment and guilt is also made by Champlin (1976: 85).

with for the purposes of this book does permit us to reject the definitional response on these grounds, it would be an act of bad faith for me to rely on this fact.

Fortunately for me, however, there is a second reply to the definitional response, and this one seems to me to be decisive. The response is that a proponent of the punishing the innocent objection can continue to maintain that the act-utilitarian solution has an unacceptable implication and simply use different words to refer to the unacceptable implication. Rather than saying that the solution would entail that the state should sometimes "punish" innocent people, for example, the objection could instead maintain that the justification the solution provides for fining, imprisoning, and executing guilty people as forms of punishment would equally justify fining, imprisoning, and executing some innocent people as forms of "pseudopunishment" or "quasi-punishment." It should be clear that anyone who finds this implication objectionable on the first description will find it equally objectionable on the second description. The problem raised by the punishing the innocent objection, that is, is not a problem about what to call the treatment of innocent people that the act-utilitarian solution would sometimes justify; rather, it is a problem with the treatment itself. And so, it should also be clear that the definitional response to the punishing the innocent objection is ultimately unacceptable.[9]

2.1.2.2 *The Denying the Implication Response*

Let us now assume that the defender of the act-utilitarian solution cannot avoid the implication that innocent people should be punished simply by appealing to the definition of the word punishment. This does not mean that she cannot avoid the implication at all. It simply means that she must provide some other reason for believing that she can. And given that she is an act-utilitarian, there is only one reason that she can give: she can try to show that the punishing the innocent objection overlooks certain negative consequences that arise from the act of (pseudo)punishing[10] an innocent person, and that once these further consequences are taken into account, it will turn out that the right thing to do, on act-utilitarian grounds, is not to punish innocent people.

The most common argument by far for this denying the implication response, appeals to the costs involved in trying to keep the practice a secret. Lyons, for example, argues that there will always be some chance

[9] Ten does recognize that this is also a good enough response (1987: 17). See also Brandt (1959: 495), Armstrong (1961: 153–4), and Kleinig (1973: 12–13).

[10] For simplicity, I will continue to speak as if it is unproblematic to refer to the act as an act of "punishing" an innocent person. To the extent that this proves mistaken, the reader can silently insert "pseudo-" or "quasi-" before each instance of the term.

of detection and that if the deliberate punishment of the innocent were discovered, "it would break down all trust and respect for law" and "the evil of such a breakdown would outweigh any good that might be obtained by punishing the merely supposedly guilty" (1974: 346, 347).[11] Once this further fact is taken into account, this argument maintains, the act-utilitarian solution to the problem of punishment does not have the objectionable implication it initially seems to have. In every case in which we might be tempted to deliberately punish an innocent person, that is, the long-term expected utility of doing so will turn out to be less than the long-term expected utility of not doing so. Call this the "secrecy argument."

There are three reasons to reject the secrecy argument. The first is that even if we take into account the further negative consequences that the argument urges us to consider, the result of the utility calculations will still clearly favor the conclusion that an innocent person should be punished (or pseudopunished) in at least some actual cases. In the first place, in many cases the probability of detection will be extremely small. What must be detected, after all, is not simply that the person sentenced was innocent, but that the judge who sentenced him knew this. In many cases, it might be possible to destroy the only material evidence of the person's innocence, but even if the person was later proved to be innocent, this would provide no evidence that the judge knew this ahead of time. Indeed, in many cases, it is difficult to imagine what could count even in principle as convincing evidence that the judge knew the person she was punishing was innocent at the time that she punished him. In addition, even if the deliberate punishment of an innocent person is uncovered, it is unreasonable to suppose that this fact in itself will have widespread consequences. Isolated examples of police and judicial misconduct are uncovered all the time, after all, and it is implausible to maintain that each such instance generates a noticeable decline in respect for the law generally. If the discovery that a single judge in a single case punished an innocent person because she was taking a bribe does virtually nothing to change people's general attitudes toward the law, for example, then the discovery that a single judge in a single case punished an innocent person to prevent a riot would presumably do even less.

The second reason to reject the secrecy argument is that it applies to only some of the scenarios raised by the punishing the innocent objection. It is relevant to the case in which the state fabricates evidence to frame an innocent person, for example, and, to a lesser extent, to cases in which the public already believes in the defendant's guilt without the need for such deception. But the response is completely irrelevant to at least one important case of punishing the innocent: that of vicarious

[11] See also, e.g., Hare (1986: 220).

punishment. Suppose, for example, that a parent is convicted of committing a particular offense in order to benefit his child. Perhaps he steals something to give to the child, or forges a document on the child's behalf, or sabotages the performance of the child's rivals in some athletic, artistic, or academic competition. Punishing the parent alone may be a relatively ineffective deterrent. Most parents are already willing to incur substantial risks of harm on behalf of their children, and so those parents who would consider committing similar offenses on behalf of their children are unlikely to be deterred by the worry that it might end up costing them something. But suppose that a judge, instead, inflicted some harm not only on the offending parent, but on the child as well (perhaps by confiscating something of value from the child in the first case, or prohibiting the child from entering further competitions in the last case). It seems reasonable to suppose that this further punishment would have a strong additional deterrent effect while requiring the infliction of only a relatively minor additional amount of harm on the children of those who persist in offending.

In this case of punishing the innocent, it is important to recognize, there is no need to keep the practice a secret and thus no basis for accepting the secrecy argument. Indeed, punishing the innocent in this case produces the significant increase in total utility that it does precisely because it is done in the open. And while offenses that a parent commits to benefit his child may provide one of the most vivid instances of this problem with the secrecy argument, there can be many others as well. Offenses committed to benefit a spouse, friend, parent, lover, political party, business organization, and so on could all be treated in the same manner, and even in cases in which the offense is committed for straightforwardly self-interested reasons, inflicting some additional punishment on an offender's loved ones might well generate greater benefit from deterrence than harm in the innocents it would target.[12] The secrecy argument has nothing to say in such cases; thus, even if it does provide a reason to accept the denying the implication response in some cases of punishing the innocent, it cannot justify accepting it in all such cases. Therefore, it cannot protect the act-utilitarian solution from the punishing the innocent objection.

There is a third reason to be dissatisfied with the secrecy argument in particular and with the strategy of denying the implication by appealing

[12] It might be objected that such offenders are too selfish to be deterred by the prospect of harm to other members of their family, but while this might be true in some cases, it seems plausible to suppose that in many other cases it is not. Courts typically allow an offender's family to put up the collateral when the offender posts bail, for example, even in cases where the offender is accused of committing a selfish offense, and this procedure generally prevents the offender from fleeing because generally he does not want his family to suffer the loss of the collateral.

to overlooked negative consequences in general. This reason, however, depends on first rejecting one final way in which a defender of the act-utilitarian solution might respond to the punishing the innocent objection. Before presenting this third problem with the secrecy argument, therefore, let us turn to the subject of this third approach.

2.1.2.3 *The Accepting the Implication Response*

Suppose that an act-utilitarian conceded that her solution cannot evade the punishing the innocent objection either by appealing to the meaning of the word punishment or by denying that her solution entails the moral permissibility of sometimes punishing innocent people. If this were so, one final option would remain. She could simply bite the bullet and accept the implication. This response is most famously associated with J. J. C. Smart. Although he is careful to note that he is "not happy" to admit that the act-utilitarian solution has this implication, Smart responds to the claim made by the objection by conceding that in order "to be consistent," he must accept it (1973: 71, 72).[13] Simply saying that one is willing to bite the bullet at this point, however, is not the same thing as providing an argument for the claim that biting the bullet is a reasonable response to the objection. We might admire Smart for having the courage of his convictions but we must still ask, more importantly, whether biting the bullet could be consistent with our own convictions. To consider this question carefully, we must first make sure that we are clear about why we are supposed to find the implication unacceptable in the first place. And once we do this, it should become clear that simply accepting the implication is unacceptable for two very different reasons. One reason involves an appeal to our moral intuitions. It aims to show that simply biting the bullet makes the act-utilitarian solution fail a reasonable application of the foundational test that any solution to the problem of punishment must pass. And while I suspect that this first reason will be sufficient for all but those who are strongly wedded to consequentialist moral theory in general, I must concede that it is unsuccessful in the face of those who are willing to countenance such a deep assault on their intuitions. But the other reason for rejecting the biting the bullet response has nothing to do with our moral intuitions; instead, it appeals to the claim that biting the bullet would make the act-utilitarian solution fail the entailment test. And on this second account, at least, biting the bullet can be seen to be unacceptable even on terms that the pure consequentialist can and must accept.

[13] As Primoratz notes, this is the basis for Daniel Dennett's term "to outsmart," meaning "to embrace the conclusion of one's opponent's *reductio ad absurdum* argument," as in "They thought they had me, but I outsmarted them. I agreed that it *was* sometimes just to hang an innocent man" [1989a: 54; Primoratz also notes that Bentham, too, responds to the objection by accepting the implication (1989a: 43)].

2.1.2.3.1 THE COUNTERINTUITIVENESS PROBLEM.

The first reason for rejecting the biting the bullet response, and the one most frequently appealed to in the literature on punishment, is that the implication is so strongly counterintuitive. This provides a reason for concluding that the act-utilitarian solution fails the foundational test: the moral theory that serves as the solution's foundation has an implication that is so morally unacceptable as to show that the theory itself is false. And if the theory itself is false, then it doesn't matter whether or not it would entail the permissibility of punishment if it were true.

Whether or not the counterintuitive problem suffices to overturn the biting the bullet response depends on one's moral intuitions. And although it is difficult to know how to argue about what intuitions people will have about what cases, I am inclined to suppose that for many readers, this version of the problem with the biting the bullet response will prove to be sufficient. The implication of the act-utilitarian solution, it is important to remember, is not simply that it would be morally *permissible* for the state to punish an innocent person, but that it would be morally *required* for the state to do so. In addition, the solution implies that it would be morally impermissible for the state's chosen scapegoat to refuse to go along with the charade in those cases involving a person being framed. Finally, and for many people most importantly, it is important to emphasize that what makes the act-utilitarian solution so deeply counterintuitive here is not simply its implication that there might be *some* cases in which the state should deliberately harm an innocent person. Many people will agree that if the stakes are high enough, this can be permissible. If the state had to imprison an innocent person to prevent the annihilation of all life on earth, for example, then most (though not all) people would be willing to concede the state's moral right to do so. What is so deeply disturbing about the act-utilitarian position, then, is not that it implies that it could ever be acceptable to intentionally harm an innocent person, but rather that in determining in which cases it would be permissible, it gives absolutely no independent weight to the fact that the person being harmed is innocent. The amount of total utility it would take to justify imposing ten units of harm on an innocent person must be precisely the same as the total amount it would take to justify imposing it on a guilty person. Commonsense morality might permit the state to deliberately harm an innocent person as a means of avoiding a catastrophe, but the act-utilitarian solution will insist that the state must deliberately harm an innocent person every time this will produce at least a little more utility overall than not doing so. This is the reason that biting the bullet is ultimately so counterintuitive.[14] And this is one reason that

[14] McCloskey (1967: 93) makes this point clear. On top of this, as Sverdlik emphasizes (1988: 193), deliberately punishing the innocent also (often) involves the state's acting

all but the most dedicated consequentialists should refuse to accept the biting the bullet response to the punishing the innocent objection.[15]

There is a second, less familiar, but much deeper reason to reject the biting the bullet response. This reason corresponds to the entailment test rather than the foundational test. Before we turn to it, however, it is worth noting that if one accepts the reason for rejecting the biting the bullet response provided by the counterintuitiveness problem, this in turn provides one further reason to reject the secrecy argument for the denying the implication response considered in the previous section. This reason has to do with the point of reductio ad absurdum objections that are used to test our moral intuitions more generally and is perhaps best explained by an example.[16] So, consider the following suffering-minimization version of act-consequentialism: the morally right act in any situation is the act that will result in reducing of suffering to a minimum. And consider the following reductio ad absurdum objection to that position: it implies that, if you had the ability to do so, the morally best action would be for you to painlessly and instantaneously destroy all sentient life in the universe. This implication is clearly intolerable. But now, suppose that a defender of the suffering-minimization theory attempted to respond to the objection by saying the following: yes, it is true that my theory implies that this would be the right thing to do if you had the ability to do so, but this is no problem for my theory because you don't, in fact, have the ability to do so. If you follow my theory given the abilities that you actually do have, it will recommend that you behave in a manner that generally conforms to your moral intuitions.

This response to this particular objection to the suffering-minimization theory is unsatisfactory because it misses the point of the objection. The point of the objection is not to prove that the theory does, in fact, direct

deceptively – yet another reason to reject the claim that doing so would be permissible, let alone obligatory.

[15] Philips attempts to justify biting the bullet here by pointing out that inadvertent punishment of the innocent will inevitably occur in any legal system that includes punishment. Virtually everyone agrees that this fact alone is not a strong argument against punishment in general. But if it is so, Philips argues, then the punishing the innocent implication is not a problem for utilitarian justifications of punishment in particular. After all, he asks, "is there really an important moral difference between punishing people we believe to be innocent and adopting a policy that we know has as a necessary consequence that innocent people will be punished?" (1985: 389). The answer to Philips's question, however, and the reason for rejecting his suggestion, is that when we punish those we believe to be innocent, we intentionally harm them for the benefit of others, whereas when we adopt a policy that we know inadvertently punishes the innocent as a consequence, we merely foresee that they will be harmed. The former but not the latter objectionably treats the innocent as a mere means of promoting the overall good, and this is why the former but not the latter strikes most people as strongly objectionable.

[16] The example is borrowed from Parfit (1986: 150n).

you to do a deeply immoral act. Rather, the point is to demonstrate that you should not accept the theory because it entails a judgment that you do not accept. And the judgment that the theory entails is unacceptable because it tells you that the only reason that you should not painlessly and instantaneously destroy all sentient life in the universe is that you lack the ability to do so. Indeed, the theory entails that you should view this inability with deep regret, since if you had that ability, it would enable you to do what the theory determines to be the morally best of all possible acts. But you do not, in fact, accept these judgments, and so you cannot, on pain of inconsistency, accept the theory that entails them.

This lesson about the way that reductio ad absurdum objections relate to our moral intuitions should be clear in such an extreme example. But precisely the same reasoning applies with precisely the same force to the secrecy argument for the denying the implication response to the act-utilitarian solution to the problem of punishment. Even if the state does lack the ability to keep its secrets absolutely secure, for example, we can easily imagine a hypothetical situation in which it has this ability. Although the scenario itself is clearly imaginary, the response it generates in us is real: the judgment that even under such circumstances, it would still be objectionable for the state to deliberately punish an innocent person simply because this would produce an increase in total utility. The secrecy argument is by its very nature unable to account for this fact.[17] Indeed, it is committed to precisely the opposite assessment: that it is deeply regrettable that the state lacks the ability to keep such a secret at no cost to itself, because if it had the ability to do so, then it would rightly use it to punish innocent people whenever this would increase overall social utility. As long as you are unwilling to accept this implication, you must refuse to accept the theory that logically entails it. And so, the claim that punishing the innocent always turns out in fact to be impractical is irrelevant to assessing the merits of the punishing the innocent objection when the objection is used to show that the act-utilitarian solution fails the foundational test. What matters is only that the act-utilitarian is committed to the conclusion that it would be the right thing to do if it were practical.

2.1.2.3.2 THE INCONSISTENCY PROBLEM. The deeply counterintuitive nature of the punishing the innocent implication should suffice to motivate most people to reject the biting the bullet response. But for those who are willing to accept the counterintuitiveness of the results, there is a second and deeper problem. This second reason for rejecting the biting the bullet response has nothing to do with our moral intuitions. Instead, it has to do with the nature of punishment. For if the defender of the act-utilitarian solution concedes that the reason she has given for

[17] This response is endorsed by, e.g., Benn (1967: 11–12) and McCloskey (1963: 599).

deliberately harming guilty people by, for example, incarcerating and executing them is also a reason for sometimes deliberately harming innocent people by, for example, incarcerating and executing them, then her position does not justify punishment in the first place, even of those who are guilty. It justifies deliberately inflicting harm on many people who are guilty, of course, but this is not the same thing as justifying their punishment. And since this is so, accepting the implication that act-utilitarianism would justify punishing innocent people would render the act-utilitarian solution unable to pass the entailment test. Even if the implication does not make a proponent of the act-utilitarian moral theory reject it as a moral theory, that is, it must still make her reject the act-utilitarian solution to the problem of punishment.

Since the claim that simply biting the bullet renders the act-utilitarian unable to justify punishment may seem a bit unclear,[18] an analogy may be of use. So, consider first an extreme case. Suppose that a defender of punishment sought, in particular, to justify the claim that capital punishment is morally permissible in the case of murder. And suppose that, in support of this claim, he offered an argument demonstrating that it would be morally permissible to kill every human being on earth. Perhaps the argument is grounded in foundational principles concerning the rights of animals and the environment that are threatened by humanity. One reason to reject his argument, of course, is that most people would find its implications morally repugnant. But a distinct reason to reject it would be this: there is a difference between saying that all murderers should be killed and saying that all murderers should be killed *because they are guilty of murder*. This biocentric argument for killing all human beings would justify the former claim but not the latter, and it is the latter that one must justify in order to justify capital punishment. As we saw in Section 1.1.5, punishment is essentially retributive in nature. It involves not simply harming an offender, but harming an offender *because* she has committed an offense. A biocentric argument in defense of killing all human beings would, by its very nature, be incapable of justifying this claim. And so we would, at least in this extreme case, clearly be entitled to say that even if the argument succeeded in justifying the claim that all murderers should be killed, it could not succeed in justifying the claim that all (or even any) murderers should be *punished*. The argument, even if it were accepted

[18] Indeed, some writers have explicitly rejected this claim. Dimock, for example, although defending punishment on retributivist grounds, maintains that the problem of punishing the innocent "may speak against [utilitarianism] as a moral theory, but not against utilitarian theories of punishment particularly" (1997: 43). But, as I argue here, the problem does count against utilitarian theories of punishment in particular because punishment involves treating the difference between legal guilt and innocence as morally relevant in a way that is inconsistent with punishing innocent people.

as sound, could not be accepted as a solution to the problem of (capital) punishment.

But now imagine that the scope of this biocentric argument was reduced, so that it entailed only that we should kill all people who are murderers and, say, 90 percent of those who are not. Or 85 percent or 80 percent. Suppose, at last, that we arrived at the position that for the good of the earth and all nonhuman life on it, we should execute every murderer and a very small handful of innocent people as well. The implications of this greatly modified version of the biocentric argument would be substantially less horrifying than those of the argument for exterminating all of humanity. But it would be just as true of this version of the argument as it was of the much stronger original version that what it would be justifying would not be capital punishment. Punishment, after all, is a practice in which the fact that one person is guilty and another is innocent is in itself treated as a reason to harm the former but not the latter. And whatever this biocentric argument would be justifying, it is not this. Yet, the reasons for saying that the biocentric argument would not justify capital punishment apply equally to the act-utilitarian attempt to justify punishment generally if one agrees that the act-utilitarian argument would justify punishing at least some innocent people for the good of society. In both cases, the argument would justify harming many people who are guilty, but it would not justify drawing the line between guilt and innocence that the institution of punishment draws. This is not to say, of course, that the implication proves that the act-utilitarian cannot justify the claim that we should harm many guilty people and some innocent ones. Rather, it is to say that even if he succeeds in justifying this claim, this cannot justify punishing people for breaking the law.

I conclude that proponents of the act-utilitarian solution to the problem of punishment cannot avoid the punishing the innocent implication and that, whether or not they are committed to a consequentialist moral theory, they cannot accept it either. While the objection does not necessarily provide a sufficient reason to reject act-utilitarianism itself, it does provide a sufficient reason to reject the act-utilitarian solution to the problem of punishment. For most people, it will show that the solution fails the foundational test. For all people, it will show that the solution fails the entailment test.

2.1.3 The Not Punishing the Guilty Objection

Although the punishing the innocent objection is the most common objection to consequentialist justifications of punishment in general and the act-utilitarian justification in particular, it is not the only one. A second important objection is, in effect, the mirror image of the first. For just as there are some cases in which the most utility is produced by

deliberately punishing an innocent person, so are there cases in which the most utility is produced by deliberately not punishing a guilty person. And this fact provides a second reason to reject the act-utilitarian solution to the problem of punishment.[19] One can imagine, to begin with, cases where the offender is so widely beloved that the anguish caused to all those who would hate to see him suffer would outweigh the benefits that would accrue from punishing him. Or one can imagine cases in which a particular offender could contribute more to the overall good in other ways than by being punished (by agreeing to leave his medically unique body to science, bequeathing his money to worthy causes, etc.). And one can imagine cases in which the state could achieve all the deterrent benefits of punishment by pretending to punish him. In all of these cases, the defender of punishment is committed to the view that it is permissible for the state to punish the offender, but in none of them can the act-utilitarian solution provide a justification for this claim.

Finally, and most importantly, it must be remembered that not punishing an offender does not require that the state do nothing at all to the offender. There may be many cases, for example, in which compelling an offender to compensate her victim would generate a lot of deterrence and in which the extra deterrence that would be added by punishing her would not outweigh the harms involved in punishing her. The same could be said of many other nonpunitive responses to violations of the law, including intensive supervision, mandatory drug and alcohol treatment, therapy, and so on. A district judge in Colorado, for example, recently changed the sentence of a young woman convicted of vehicular homicide from four years in prison to five years of probation with the condition that she speak to students about the dangers of drinking and driving [Anas (2005: 2A)]. It seems eminently plausible that the new sentence will produce more overall good than the old one, but the old sentence was clearly punitive, while the new one is not. Indeed, there may be many cases in which the mere act of publicly convicting an offender would generate enough shame to deter most people from committing a comparable offense and in which, again, the extra deterrence added by punishment would not outweigh the harms that punishment would involve. Brandt, for example, cites a former judge who reported that "of the hundred or so persons he had sentenced for embezzlement in his years as a senior judge of the Federal Court in Boston, he had given everyone a suspended sentence and believed that the effects of the crime on career, publicly known, are deterrence enough" (1995: 83). For a defender of punishment, the line between those who break the law and those who do not is treated as morally relevant. If punishment is justified,

[19] This has also been noted by some writers, e.g., Primoratz (1986: 41–3) and Gavison (1991: 352).

then breaking a just and reasonable law is sufficient to allow the state to punish the person for doing so. The not punishing the guilty objection shows that the act-utilitarian solution is unable to account for this. For many people, of course, this result will be far less counterintuitive than the one pointed to by the punishing the innocent objection. But even if the not punishing the guilty objection does not convince them that the act-utilitarian solution fails the foundational test, they will still have to concede that it fails the entailment test. And that is enough to show that it fails as a solution to the problem of punishment.

2.1.4 The Disproportionate Punishment Objection

Punishment involves drawing a line between those who break the law and those who do not and treating that line as morally relevant. I have argued that act-utilitarianism is unable to justify drawing this line and that the act-utilitarian solution to the problem of punishment is therefore unsuccessful, regardless of whether or not one accepts act-utilitarianism as a moral theory. But let us now suppose that I have been mistaken, and that under act-utilitarianism it is permissible to punish all and only those who break the law. Even if we thereby allow that the act-utilitarian solution passes the entailment test, there remain further reasons to reject it. For there are other reasons to believe that it fails a reasonable application of the foundational test. These reasons arise from implications that the act-utilitarian solution has in a variety of other cases.

One set of further implications arises from cases in which, even if act-utilitarianism renders it permissible to punish people for breaking the law, it cannot render permissible an intuitively appropriate amount of punishment.[20] In particular, most people strongly believe that the severity of punishment should be at least roughly proportional to the offense: that trivial offenses should mandate minor punishments, while serious offenses should require severe punishments. The disproportionate punishment objection maintains that the act-utilitarian solution cannot account for these central judgments. In some cases the solution will justify an unacceptably severe punishment, and in others it will justify an unacceptably mild one.[21]

[20] The objection is endorsed, in one form or another, by a number of writers, including Hawkins (1944: 14), Armstrong (1961: 152), Hampton (1984: 126), Primoratz (1989a: 37–8), Braithwaite and Pettit (1990: 46), Gavison (1991: 356), and Cragg (1992: 47–8).

[21] In addition to the two versions of the objection developed here, it is worth noting that the same problem can arise in cases where one group of people is more easily deterred than another. If, for example, women would be optimally deterred from committing a particular offense by the prospect of a three-year sentence while men would be similarly deterred by the prospect of a five-year sentence, then the act-utilitarian solution would seem to support punishing men more than women for the same offenses with the result

One way for the act-utilitarian to respond to the disproportionate punishment objection is simply to bite the bullet. This does not generate as great a difficulty as it does in the punishing the innocent objection, since, as I argued in that case, to admit the implication is in effect to abandon punishment altogether. But this will strike most people as morally unacceptable nonetheless. Most defenders of punishment seek not only to justify the right to punish, but to justify the right to punish that is constrained by a principle of proportionality. Unless the act-utilitarian can find a way to avoid these further implications, therefore, the disproportionate punishment objection should provide a further reason for most people who believe in punishment to reject the act-utilitarian solution to the problem of punishment. The question, then, is whether the act-utilitarian can avoid the implication in the first place.

2.1.4.1 The Too Much Punishment Version

The most conspicuous version of the disproportionate punishment objection focuses on cases in which act-utilitarianism would justify punishments that are intuitively excessive. When the penalty for minor driving offenses (speeding, illegally changing lanes, etc.) is small (e.g., a twenty-five-dollar fine), for example, many people will persist in breaking the law. They believe that there is a significant chance that they will not be caught, and recognize that even if they do get caught, the punishment will be modest. If, on the other hand, the punishment for such offenses is a swift, certain, and painful execution, then virtually everyone will obey the law. Since carrying out such clearly disproportionate sentences would do a better job of promoting the public good, it follows that if the act-utilitarian is committed to punishment, then at least in some cases she is also committed to excessive punishment. And since this implication seems plainly unacceptable, it provides a reason to conclude that the act-utilitarian solution fails a reasonable application of the foundational test even if it passes the entailment test.

Some writers on punishment, including some who are hostile to consequentialism in general, are unimpressed by the disproportionate punishment objection. Nino, for example, who ultimately rejects the consequentialist approach for other reasons, nonetheless characterizes this particular objection to it as "clearly absurd." Since "the death of one person is worse than a congested traffic flow," he argues, the consequentialist would not be committed to such an implication in the first place (1983: 95–6; 1991: 264–5).

that either men would get more punishment than they deserve or women would get less than they deserve [Smilansky (1992: 124–6) develops this worry in the context of different socioeconomic groups on the plausible assumption that the poor might require a greater threat of harm than the wealthy to produce a comparable level of deterrence].

This response to the too much punishment version of the disproportionate punishment objection is initially appealing. Certainly it is far worse for one person to die than for one person to be stuck in traffic. But that is not the relevant trade-off. The relevant trade-off is between one person dying and many people being stuck in traffic. And once this is taken into account, things begin to look much worse for the act-utilitarian position.

At first, it may seem that this attempt to defend the too much punishment version of the disproportionate punishment objection rests on an uncharitable assumption about the costs and benefits of preventing minor traffic infractions. But, in fact, this response can be sustained purely by appealing to the assessments of such considerations that most people already make and accept all the time. Consider, for example, the following scenario. The speed limit in a given state is currently sixty-five miles per hour. A study determines that if it were reduced to fifty-five miles per hour, ten fewer innocent people would be killed in auto crashes each year. But it also determines that this would be more inconvenient for tens of thousands of drivers and recommends against lowering the speed limit. This is a very common occurrence. We accept the foreseeable deaths of a few innocent people as the price we are willing to pay for more convenience on the highway. Certainly the act-utilitarian can't complain about this.

Now consider a second case. The speed limit in another state is currently fifty-five miles per hour, but it is largely ignored. The state tickets people for speeding when it catches them, but many people do not get caught and most who do can easily afford to pay the tickets. A study determines that 50,000 people break the speed limit in the state each year, resulting in many accidents and deaths of innocent people. The study also determines that if the state instituted a mandatory death penalty for speeding, compliance would be virtually 100 percent. Still, it estimates that a handful of people, say five a year, would still end up breaking the law. There is no reason to doubt that, in the first case, utilitarian calculations will often favor accepting a few more deaths in order to get a much smaller benefit but for many people. And I see no reason to think that the second case would be any different. Executing five people for speeding each year would produce better results overall. So, the act-utilitarian would be committed to endorsing it. This result seems unacceptable to most people and, to the extent that it is, it provides another reason to conclude that the act-utilitarian solution fails the foundational test and thus fails to solve the problem of punishment.

2.1.4.2 The Too Little Punishment Version

The prospect of executing people for relatively minor offenses is the most disturbing version of the disproportionate punishment objection, but it is

important to note that the act-utilitarian solution is subject to the same objection from the opposite direction. There will be cases, that is, where, even if the solution justifies punishing lawbreakers, it will only justify inflicting a minor punishment for a serious offense. And most people who believe in punishment will also find this sufficiently counterintuitive to warrant rejecting the act-utilitarian solution. Such cases will arise because the justification for each act of punishment, on the act-utilitarian account, rests on the claim that punishing an offender will deter people from committing offenses in the future. Since the suffering of the offender is admitted to be a bad thing in itself, it follows that the act-utilitarian will not wish to impose any more suffering on the offender than is needed to deter others. And in some cases, the amount of suffering needed to generate the optimal level of deterrence will be considerably less than the amount that the offender would inflict on his victim.

Consider, for example, a certain category of arson: that which causes over $1 million worth of damage to property but no injury or loss of life. Suppose that the vast majority of people who engage in this sort of arson do so simply because they find it entertaining. They would not be deterred from burning down a $1 million house by a $10 fine, because they would happily pay $10 for the enjoyment they would receive from it, but they would be fully deterred from burning down such a house by the prospect of paying a $500 fine. For that amount of money, they could buy more entertainment in other forms. Suppose that this accounts for all but a handful of would-be arsonists. Suppose, furthermore, that virtually all other potential arsonists are virtually undeterrable. Their desire to burn buildings is so strong, that they will try to do so even under the threat of death. The handful of arsonists who lie between these two extremes also have an extremely powerful desire to engage in arson, but they could be deterred by the prospect of execution. And now, suppose that the state is attempting to determine, on act-utilitarian grounds, the appropriate amount of punishment for arson in such cases.

One possibility is that it merely imposes a $500 fine. This will be enough to deter virtually everyone. A second is that it impose a punishment that is more severe than a $500 fine but less severe than death. But no punishment in this range will deter anyone who is not already deterred by the $500 fine. Since any punishment in this range would thus cause more suffering to the offenders and produce no increase in deterrence, it must be rejected. The third option is to impose the death penalty in such cases. This would deter a few more people, but would result in many more deaths and produce only a minor reduction in arson. In this case, then, the act-utilitarian can justify only the $500 fine for committing an offense that causes its victim to lose a $1 million house. And, once again, to the extent that the resulting punishment seems intuitively disproportionate, the argument that justifies it fails a reasonable

application of the foundational test and thus fails to provide a satisfactory solution to the problem of punishment.

2.1.5 The No Excuses Objection

A further set of problems for the act-utilitarian solution arises from cases involving mitigating circumstances. Suppose, for example, that Larry punches Moe in the nose, Curly punches Shemp in the nose, and the two assaults are identical in all respects except that Moe provoked Larry, while Shemp did nothing to provoke Curly. Now, the fact that Larry was provoked into punching Moe is certainly not a justification for punching him. It does not mean that Larry had the right to do what he did. And so, if a solution to the problem of punishment is to succeed, it should account for why it would be permissible to punish both Larry and Curly. But while it is clear that Larry had no justification for punching Moe, it seems plausible to suppose that Larry's provocation by Moe gives Larry a mitigating excuse. This means that, if punishment is merited in cases of physical assault, then while Larry and Curly should both be punished for what they did, Curly's punishment should be greater. This is the way our legal system operates, and this particular feature of the system seems to be relatively uncontroversial. In addition, the claim that the degree of punishment should differ in the two cases seems well supported by the belief that punishment, if it is justified at all, should vary with the degree of responsibility and the belief that, in some important respect, Larry is less responsible (though responsible nonetheless) than Curly. So, a justification of punishment would have a significant problem if it could not account for this important claim.

And yet, not only is the act-utilitarian solution unable to account for this claim, but it in fact entails precisely the opposite result: that Larry should receive a greater punishment than Curly and that, more generally, people who commit crimes under duress or in the heat of passion should receive greater sentences than those who do not.[22] This problem arises from the same feature of act-utilitarianism that gives rise to the too little punishment version of the disproportionate punishment objection. The justification for each act of punishment, on the act-utilitarian account, rests on the claim that a given punishment of a given offender will deter people from committing crimes in the future. Since the suffering of the offender in itself is admitted to be bad (that is, it is undesirable in itself and desirable overall only because of its deterrent effect), it follows that the act-utilitarian will not wish to impose any more suffering on the offender than is needed to deter others. If putting a particular person in jail for three years for committing a particular offense will deter most people from committing that offense in

[22] This objection is pressed by, e.g., Ten (1987: 147–8) and Primoratz (1989a: 38).

the future, for example, and if dramatically increasing the sentence will not increase deterrence sufficiently to compensate for the increased suffering of the offender, then the act-utilitarian will set the punishment for that offender at three years and no more.

Let us suppose that this is the case with the crime of punching a person in the nose without provocation: for the vast majority of people, the prospect of three years in jail is sufficient to deter them from punching other people in the nose without provocation, and the costs of increasing the punishment beyond three years would, on the whole, outweigh the benefits. In that case, putting Curly in jail for three years for his unprovoked attack on Shemp will be the right thing to do on act-utilitarian grounds: punishing Curly in this manner will have better overall consequences than will any available alternative. But now consider the fact that when a person is provoked to anger by another, the thought of a penalty that would, under normal circumstances, be sufficient to deter him will no longer be sufficient. An angry person will be more likely to focus on the immediate satisfaction he will get from punching his provoker and less likely to focus on the long-term cost he will incur by being punished for it. If, on the other hand, a person under such provocation knew that people who punch others in the nose after being provoked receive a much greater sentence (maybe ten years in prison, or twenty, or life), he would be much more likely to be deterred. The result, then, is that, on act-utilitarian grounds, Larry should receive a significantly greater punishment than Curly, even though it is Larry, and not Curly, who has what intuitively should count as a mitigating excuse for what he did. Since this implication of the act-utilitarian solution is intuitively very difficult to believe, it again provides reason to conclude that the act-utilitarian solution fails a reasonable application of the foundational test and thus fails as a solution to the problem of punishment.[23]

How might a defender of the act-utilitarian solution respond to this objection? One possibility, again, would be to admit that the solution has this implication and agree to accept it. This, at least on some interpretations, is Bentham's view [e.g., Ten (1987: 148)].[24] For most people,

[23] The objection can also be raised in the case of complete excuses such as insanity. As Primoratz points out (1989a: 38–41), we can't account for the merits of the insanity defense without appealing to such nonconsequentialist considerations as responsibility, just deserts, and so on. If all we care about is producing the optimal amount of deterrence, then punishing the insane would be better than not punishing them, since it would deter people who might otherwise have committed an offense with the plan of faking an insanity defense.

[24] See also Brandt (1959: 492–3), who argues (although in the context of a defense of rule-utilitarianism) that allowing for excuses is at most contingently justified, suggesting that if the elimination of excuses produced better consequences, there would be no further reason to accept them.

however, this is surely an unacceptable response: in seeking a solution to the problem of punishment, they want to justify not merely a right to punish, but a right to punish that is consistent with some of their other firmest convictions. The alternative is to attempt to establish that the act-utilitarian solution does not have this implication. But the attempts that have been made to establish this are unpersuasive. Brandt, for example, argues that "people who commit impulsive crimes, in the heat of anger, do not give thought to legal penalties; they would not be deterred by a stricter law" (1959: 493). But if people are incapable of being deterred from committing impulsive crimes, then not only is there no utilitarian basis for giving a stricter sentence to someone who commits one, there is no utilitarian basis for giving such a person any sentence at all. It is difficult to see how the act-utilitarian solution can avoid this implication; thus, it is difficult to see how it can avoid succumbing to yet another reason to conclude that it fails a reasonable application of the foundational test that a successful solution to the problem of punishment must pass.

2.1.6 The Treating People as a Means Objection

I have argued thus far that the act-utilitarian solution fails to justify the morally relevant line that punishment draws between the guilty and the innocent. I have also argued that even if it did draw the line in the appropriate place, it would still fail to justify the morally relevant lines within the set of people who should be punished that distinguish those who merit less punishment from those who merit more. If the law treated people in the way that act-utilitarianism entails that they should be treated, the result would look very different from the practice that defenders of punishment seek to defend. But now I want to assume that I have been mistaken. Let us suppose that in every instance, the greatest utility is brought about by not punishing innocent people and by punishing guilty people in a manner that seems proportionate to their offenses. There remains one final problem with the act-utilitarian solution. It is a problem that a true utilitarian will simply accept as the price of accepting utilitarianism, and so once more provides no reason to suspect that the solution fails to pass the entailment test. But it is a problem that virtually everyone else should accept as a reason to conclude that it fails the foundational test, and thus as one final reason to reject the act-utilitarian solution.

The problem is that even if the act-utilitarian solution correctly identified the right people to punish and the right amounts of punishment, it would still do so in a manner that is morally objectionable. For it justifies harming these people merely on the grounds that harming them will benefit others. This is to treat them as mere means of promoting the public good. We would never treat innocent people this way just to promote the public good. And so, we should not treat guilty people this

way either. They are still, after all, people. For the reasons developed in Sections 2.1.2 and 2.1.3, the claim that act-utilitarianism is true does not support the claim that punishment is permissible. But even if it did, it would be morally wrong to justify punishment by appealing to it.

Now, baldly stated, there might seem to be a very easy response to this objection: yes, a defender of the act-utilitarian solution might say, the solution does involve treating offenders merely as a means of promoting the common good, but they are, after all, offenders. It is one thing to use an innocent person merely as a means, they might say, but quite another to treat a guilty person this way. The claim that guilty people may be treated as means while innocent people may not could turn out to be true. But if it is true, it will have to be true for nonutilitarian reasons. It might turn out to be true, for example, because guilty people deserve to be treated this way, or have consented to be treated this way, or have forfeited their right not to be treated this way. But if a defender of punishment has to appeal to such considerations, he is at that point abandoning a utilitarian solution in favor of a nonconsequentialist one.[25] So, appealing to the mere fact that offenders are guilty cannot help the act-utilitarian escape the force of the objection.

There is a second response that a defender of the act-utilitarian solution can offer here. Yes, she might agree, the solution does entail that it permissible for the state to harm innocent people as a means of promoting the public good, but in fact the state does this all the time and nobody complains about it. Walker, for example, attempts to dismiss the objection by appealing to the following examples:

> Non-offenders too are knowingly harmed for the benefit of others. Urban redevelopment evicts families from [their] homes. Airports sacrifice the peace of a few for the convenience of many. Sufferers from some communicable diseases undergo irksome restrictions in the interests of public health. Soldiers are enlisted – not always voluntarily – to risk life and limb for [their] country. The list of examples could be much longer. The short point is that anyone who condemns deterrent or precautionary sentences on the ground that they harm offenders for the sake of others must either condemn many of the things that are done to the innocent or explain why only the guilty should be immune from being "treated as means" (1993: 54).[26]

The problem with this response to the treating people as a means objection is that it neglects the distinction between intending harm and merely foreseeing it. In all of the cases that Walker cites, the state does an

[25] This point is noted by a number of writers, including Lewis (1949: 304–5), Armstrong (1961: 152), Murphy (1973: 5–6), and Nino (1983: 96–7).

[26] Andenaes (1970: 130) appeals to some of the same examples in offering the same response to the problem.

act that causes some harm to some people and produces larger benefits for many others. Locating an airport in a particular neighborhood, for example, causes some people to suffer headaches and at the same time causes many more people to enjoy much greater benefits. But this does not involve harming the former as a means of benefiting the latter. Such a case would be one in which the state deliberately gave some people headaches – for example, as part of an involuntary form of medical research – in order to develop treatments that would then benefit others. The commonsense morality that the critic of the act-utilitarian solution is appealing to at this point can clearly accept the permissibility of the first case while rejecting the permissibility of the second. And the same distinction undermines every other example that Walker provides. Evictions, quarantines, and military conscription do not involve intentionally harming some people in order to benefit others. Rather, they involve intentionally doing acts that foreseeably cause some harm to some people and provide greater benefits to many others. It is perfectly consistent for the critic of the act-utilitarian solution to accept all of these practices as morally permissible while still insisting that the act-utilitarian solution impermissibly treats offenders as mere means to promoting the public good. Indeed, not only is this consistent, it helps to highlight precisely what makes the act-utilitarian defense of punishment so disturbing. For while Walker promises that his list could be made much longer, the fact is that punishment stands alone as the one instance in which the state not only does an act that predictably harms some of its citizens, but in which it acts with the explicit aim of causing harm. Punishment is utterly anomalous in this respect. This is precisely what makes punishment distinctively difficult to justify in the first place. And this is why appeals to such cases as eminent domain, curfews, quarantines, and so on cannot help to diffuse the power of the objection that justifying punishment on act-utilitarian grounds treats people in a morally unacceptable way. Earlier in this section, I argued that the act-utilitarian moral theory does not provide support for punishment. But even if it did, it is support that most defenders of the practice must reject as incompatible with their own moral values.

2.2 THE RULE-UTILITARIAN VERSION

I have argued that the act-utilitarian solution to the problem of punishment should be rejected. A proponent of the consequentialist solution in general might agree with this assessment and view it as a reason not to abandon the solution itself but simply to modify it. One way to modify the solution would be to appeal to a different criterion for what makes one consequence better than another. This would be to appeal to a nonutilitarian version of consequentialism, an approach that will be considered in Section 2.4. But a second and more familiar response is to take the

problems with the act-utilitarian solution as a reason to embrace a rule- rather than an act- version of the utilitarian solution. This position is most famously associated with Rawls's early article "Two Concepts of Rules" (1955). I will therefore structure my treatment of this position around a discussion of Rawls's article.[27]

2.2.1 The Rule-Utilitarian Solution

Rawls begins by explaining the difference between "justifying a practice as a system of *rules* to be applied and enforced, and justifying a particular *action* which falls under these rules" (1955: 5, emphasis added). Suppose, to follow Rawls's example, that a man is convicted of robbery and the state puts him in prison. In this case, as Rawls points out, we can ask two importantly distinct questions about the state's act of imprisoning the man: (1) why does the state put robbers in prison? and (2) why did the state put this particular person in prison? The first question is a question about a general practice or institution that the state has adopted, a practice that involves the implementation and enforcement of a certain set of rules. The question is: why does the state follow this practice rather than some other practice? The second question is a question about a particular act that falls under that set of rules. The question is: given that the state has adopted this particular set of rules, why did it put this particular person in prison at this time?

As Rawls then points out, the distinction between these two levels of justification enables the utilitarian to construct a defense of punishment as a practice or institution that is wholly consequentialist at its foundation, yet at the same time accounts for the sense in which individual acts of punishment are, by their very nature, retributive. So, in answer to the second question, a utilitarian of this sort will say: the state imprisoned that

[27] See also Benn and Peters (1959: 96); Brandt (1959: 494–5), Hare (1986: 218ff), Knowles (1999: 33), and, for a related argument, Wertheimer (1976). Rule-utilitarianism and rule-consequentialism more generally are arguably on the decline in the contemporary literature on consequentialism, so I have structured my discussion of the rule-utilitarian solution here around an older but undeniably classic article; however, for a good recent defense of a rule-consequentialist approach to morality, see Hooker (2000). Rosen (1997) argues that the classical utilitarians successfully avoided the punishing the innocent objection in a similar manner: "For classical utilitarianism, it was the secondary principles [the following of which was, in turn, justified by an appeal to the principle of utility] and not the principle of utility itself which were to serve as tests of practice." But while Rosen illustrates this analysis by writing that the "idea of a proportion between crimes and punishments was to be used to judge practice [because the idea of using proportionality can, in turn, be justified by appeal to the principle of utility]," he is only able to claim that "by that judgment it would follow that the punishment of the innocent would be *generally* excluded" (33), and, as we saw in Section 2.1.2.3, this is not good enough.

particular person because he was found guilty of robbing a bank. If we look at this question from the point of view of the judge, on this account, the only relevant considerations in deciding whether or not to imprison the man are essentially backward-looking: did this man, in fact, rob this bank, was robbing the bank illegal, were there any mitigating circumstances, and so on. The judge, on this account, would not make a decision based on the direct consideration of whether imprisoning the man now would have beneficial consequences in the future. So, on this account, the judge reasons as a retributivist, and his act of imprisoning the man is justified by appealing to a rule whose content is thoroughly retributivist: this man was convicted of robbing a bank, and I make it a rule to imprison people who are convicted of robbing banks.

But while the *content* of the rule that the judge appeals to in this case is entirely nonconsequentialist in nature, this in itself entails nothing about the *justification* of the rule. To ask about the rule's justification is instead to ask the first question raised earlier: why does the state put bank robbers in prison in the first place? This, as Rawls puts it, is to view punishment from the perspective of the (ideal) legislator rather than that of the (particular) judge. And at this point, the utilitarian can maintain that the rule is adopted in the first place because its adoption, at least in the context of an appropriate set of corresponding rules, "will have the consequence, in the long run, of furthering the interests of society."

The justification for individual acts of punishing offenders on this account is not, in itself, utilitarian. So, this position clearly departs from that of the act-utilitarian solution. But this position is nonetheless meant to justify such acts. So, it is meant to count as a solution to the problem of punishment. And since the acts are justified by being in accordance with a specific set of rules, and since following the rules is itself said to be justified on utilitarian grounds, the solution to the problem of punishment is still a utilitarian one. It is simply a rule-utilitarian solution rather than an act-utilitarian one. There is little question that this rule-utilitarian solution is different from the act-utilitarian solution. The question is whether it is better and, if it is, whether it is good enough to render it successful.

2.2.2 The Punishing the Innocent Objection

Rawls himself does not insist that the rule-utilitarian solution should be accepted. But he does argue that attending to the importance of the distinction between act and rule "strengthens" the utilitarian position and allows it to be developed in such a way as to be immune from at least some of the strongest objections leveled against it in its act- form. Since the most prominent objection is the punishing the innocent objection, and since Rawls's motivation in endorsing rule- over act-utilitarianism turns largely on the claim that the rule- formulation can overcome this

objection, I will begin with it. I will simply assume here what was argued for in Section 2.1.2, namely, that there are cases in which act-utilitarianism does make it permissible to deliberately punish an innocent person, and that any solution to the problem of punishment that has this implication thereby fails the entailment test (and, for most people, the foundational test as well) and thus fails as a solution to the problem for that very reason. The question then is whether or not rule-utilitarianism provides a way for the consequentialist to avoid this problem. At first, it might seem clear that if punishing the innocent is a problem for the act-utilitarian, then it will be a problem for the rule-utilitarian as well. After all, if more utility is sometimes produced by punishing innocent people, then the rules that will produce the most utility apparently would have to be rules that sometimes permit the punishment of innocent people. But Rawls argues that this is not so, and his argument for this claim forms the basis of the rule-utilitarian solution to the problem of punishment.

2.2.2.1 *Punishment versus Telishment*

Rawls's argument in response to the punishing the innocent objection turns on a claim he makes about a practice he refers to as "telishment." Rawls defines telishment as an institution "which is such that the officials set up by it have authority to arrange a trial for the condemnation of an innocent man whenever they are of the opinion that doing so would be in the best interests of society. The discretion of officials is limited, however, by the rule that they may not condemn an innocent man to undergo such an ordeal unless there is, at the time, a wave of offenses similar to that with which they charge him and telish him for" (1955: 11). Rawls then argues that if an ideal legislator were charged with choosing between punishment, on the one hand, and punishment plus telishment, on the other, he would endorse punishment without telishment and would do so on purely utilitarian grounds. Since this claim is crucial to Rawls's argument, his reasoning for it is worth quoting in its entirety:

> Once one realizes that one is involved in setting up an *institution*, one sees that the hazards [of telishment] are very great. For example, what check is there on the officials? How is one to tell whether or not their actions are authorized? How is one to limit the risks involved in allowing such systematic deception? How is one to avoid giving anything short of complete discretion to the authorities to telish anyone they like? In addition to these considerations, it is obvious that people will come to have a very different attitude towards their penal system when telishment is adjoined to it. They will be uncertain as to whether a convicted man has been punished or telished. They will wonder whether or not they should feel sorry for him. They will wonder whether the same fate won't at any time fall on them. If one pictures how such an institution would actually work, and the enormous risks involved in it, it seems clear that it would serve no useful purpose. A utilitarian justification for this institution is most unlikely. (1955: 12)

In short, punishment plus telishment would produce less overall social utility than punishment alone because of the difficulties of limiting telishment to the right cases and because of the effects that telishment would have on the population as a whole.

To argue that an ideal legislator reasoning on utilitarian grounds would choose to endorse the rule of punishment without telishment for these reasons, however, is not enough to establish that the rule-utilitarian solution can overcome the punishing the innocent objection. It must also be argued that this fact about what rule the legislator would choose suffices to justify the claim that judges in particular cases should act according to the rule. The problem of punishment, after all, is not simply the problem of explaining how it can be permissible to establish the practice of punishment. It is the problem of explaining how it can be permissible actually to engage in it. And so, the rule-utilitarian attempt to overcome the punishing the innocent objection must ultimately rest on two claims: that legislators reasoning on utilitarian grounds would endorse a system of rules on which innocents are never deliberately punished and that this would suffice to justify adhering to such rules even in cases where more good would be done by breaking them. The first is a claim about the content of the rules that would be selected, while the second is a claim about following the selected rules. Both claims must be sustained for the rule-utilitarian solution to overcome the punishing the innocent objection. But both claims should be rejected.

2.2.2.2 *The Rule Content Problem*

Let me begin with Rawls's argument for the first claim: that legislators reasoning on utilitarian grounds would endorse a system of rules on which innocents are never deliberately punished. Call this the "rule content claim." There are two problems with Rawls's defense of this claim. The first problem arises from the fact that Rawls imposes on his ideal legislators an unjustifiably narrow range of options. They are given only two choices: punishment, which by its nature involves never deliberately harming the innocent, and punishment plus deliberately harming the innocent only in cases of telishment as Rawls defines it. But there are two features of telishment as Rawls presents it that unfairly stack the utilitarian deck in favor of punishment without telishment so defined. The first feature concerns the burden of proof that must be met before an official can proceed with an act of telishment. Officials are ordered to carry out such acts "whenever they are of the opinion that doing so would be in the best interests of society" (1955: 12).[28] The second feature concerns the range of cases within which judges may choose to telish an

[28] Rawls makes the same point when he describes the case against telishment as one that holds that "no official should have discretionary power to inflict penalties *whenever he*

innocent person, the stipulation that "they may not condemn an innocent man to undergo such an ordeal unless there is, at the time, a wave of offenses similar to that with which they charge him and telish him for" (11). These features make telishment less appealing on utilitarian grounds than it would otherwise be.

The problem with the first feature of Rawls's account is that it makes telishment too easy for an official to justify. Given the legitimate utilitarian concerns about the impact of accidentally telishing innocent people in cases where this turns out not to produce more good than harm, the standard of proof that the official must meet should, on utilitarian grounds, be higher. It should, for example, tell officials to telish an innocent person only when it is clear to them beyond a reasonable doubt that this will be in the best interests of society. The problem with the second feature of Rawls's definition is that it arbitrarily excludes from telishment those cases that involve what is sometimes referred to as "vicarious punishment," such as punishing (that is, telishing) the child of a parent who committed an offense on behalf of the child. I will continue to refer to Rawls's version of telishment, which adopts the weaker standard of proof and excludes cases of vicarious punishment, as "telishment" and will refer to the version that adopts the stronger standard and permits vicarious punishment as "u-telishment," since it seems clearly superior to telishment on utilitarian grounds.

Once the distinction between telishment and u-telishment is recognized, the problem with Rawls's argument for the rule content claim becomes clear. Even if it is true that legislators deliberating on utilitarian grounds would choose punishment over punishment plus telishment, it does not follow that they would choose punishment over punishment plus u-telishment. And the very features that distinguish u-telishment from ordinary telishment strongly suggest that they would choose punishment with u-telishment over punishment without it.

The reasons for thinking this is so of the first feature can perhaps best be seen by considering other cases in which we evaluate competing rules concerning the permissibility of doing acts that might harm innocent people. If the choice were between never permitting a doctor to withdraw life support from a patient, for example, and permitting her to do so in every case in which it seemed to her at least a bit more likely than not that the patient was irreversibly comatose, it seems plausible to suppose that the costs arising from cases in which people were mistakenly allowed to die in the latter case would suffice to render the former rule preferable overall. But in this case, at least, it should be clear that the comparative judgment would provide little support for the adoption of an

thinks it for the benefit of society; for on utilitarian grounds an institution granting such power could not be justified" (1955: 8, emphasis added).

exceptionless rule forbidding passive euthanasia. Rather, it would suggest that the standard of proof should be higher, permitting doctors to discontinue life support only if, for example, it was clear beyond a reasonable doubt that the patient would never regain consciousness. The same is true, for that matter, of our choice regarding who to convict of offenses in the first place. The overall costs involved in allowing courts to convict anyone they believe to be more likely guilty than not are too high, but this does not lead us to conclude that courts should convict no one. Rather, it leads us to conclude that the standard of proof should be considerably higher. And the same goes for this first feature of Rawls's account of telishment.

The problem arising from the second feature of Rawls's definition is far more severe. For the second feature simply excludes by fiat an entire range of cases in which it is plausible to suppose that punishing an innocent person will be justified on utilitarian grounds: those cases involving vicarious punishment. As I argued in Section 2.1.2.2, for example, cases in which a parent commits an offense in order to benefit his child provide a clear opportunity to produce a significant gain in deterrence at a relatively minor cost by aiming some punishment at the child and not just at the parent. Other parents who would be tempted to break the law to benefit their own children may barely be deterred, if at all, by the risk of harm to themselves, but they will be much more powerfully deterred by the worry that their acts might ultimately harm their own children. The case of vicarious punishment undermines Rawls's argument for the rule content claim because the negative consequences he identifies in his version of telishment simply do not apply to vicarious punishment. Since there is no reason to keep the practice secret, there are no costs involved in keeping it secret and no reason to believe that judges would abuse their authority to telish in such cases any more than they would abuse their authority to punish in others. Rawls might well be right to argue that ideal legislators would select punishment over punishment plus his version of telishment, then, but he is wrong to think that this shows that they would select punishment. Instead, they would select punishment plus, at the very least, some cases of vicarious punishment of the innocent. And this is enough to undermine the rule content claim.

I have argued thus far that Rawls's telishment-based response to the punishing the innocent objection is subject to a rule content problem. Once we consider all of the options that would be available to ideal legislators ruling on utilitarian grounds, we must conclude that such legislators would in the end choose punishment plus some form of telishment over punishment without it. If I am correct, then rule-utilitarianism cannot overcome the punishing the innocent objection. But let us now suppose that I have been mistaken and that when all the relevant utility calculations are done, an exceptionless rule of only punishing the guilty

proves superior to any version of punishment plus telishment that permits the occasional telishment of the innocent. If this is so, then rule-utilitarianism will endorse the adoption of a rule by which innocent people are never punished. And this will help put the rule-utilitarian solution in a position to pass the entailment test (though, as we will see in the subsection that follows, the rule-following problem will still show that the solution fails the test). But even if this is so, the rule content problem will still undermine the rule-utilitarian response to the punishing the innocent objection for a second reason. For even if it turns out to be true that telishment is too difficult to carry out effectively in practice, this fact is too contingent[29] to account for the intuitive force of the punishing the innocent objection. To the extent that the punishing the innocent objection is taken to show that a given solution fails a reasonable application of the foundational test, that is, rather than to establish that it fails the entailment test, the objection rests on the claim that the solution in question has implications that are strongly at odds with our considered moral judgments. And, understood in this way, Rawls's telishment-based defense of the rule content claim does nothing at all to undermine the objection.

When the punishing the innocent objection is understood as resting on an appeal to intuition rather than on an appeal to the nature of punishment, its force arises not from the thought that we should not punish innocent people, but rather from the thought that we should not do so even if we could do so effectively. Even if it was easy to keep secrets, to oversee judges, to limit their discretion, and so forth, that is, we would think it wrong to authorize them to punish innocent people simply as a means of increasing overall utility. And because the rule-utilitarian defense of punishment alone over punishment with telishment rests on such contingent facts, it is by its very nature unable to account for this intuition. Indeed, it is again committed to precisely the opposite assessment: that it is regrettable that the state lacks the ability to telish effectively because, if it had the ability to do so, then it would rightly use it to deliberately telish innocent people. Most people will reject this assessment. For them, the fact that the rule-utilitarian solution entails this assessment will therefore count as a further reason to accept the punishing the innocent objection. Even if Rawls's response succeeded in helping the rule-utilitarian solution to pass the entailment test, therefore, it would still fail to help it pass a reasonable application of the foundational test. And that would be enough, for most people, to conclude that the rule-utilitarian solution to the problem of punishment is unsuccessful.

[29] As Rawls himself acknowledges, this hinges entirely on considerations about "how such an institution would *actually* work" (1955: 12, emphasis added).

2.2.2.3 *The Rule Worship Problem*

I have argued that the rule content problem prevents the rule-utilitarian solution from overcoming the punishing the innocent objection for two different reasons. But let us now assume that I have been mistaken on both counts. Let us assume that the right thing for a utilitarian legislature to do is to create a legal system in which judges have no legal discretion to deliberately punish innocent people, and let us assume that we are not bothered by the contingency of the facts that must be appealed to in making good this claim. If all of this is conceded, a second problem will now arise: why, on utilitarian grounds, should judges always follow the rules that the legislature sets down for them?[30] This problem is not limited to a rule-utilitarian justification of punishment in particular. It arises for any application of rule-utilitarianism. The mere fact that adopting a rule at one point in time might be the best thing to do from a utilitarian point of view at that point in time does not entail that following that rule at every subsequent point in time will be the best thing to do from a utilitarian point of view at those points in time. And if one now finds that one can produce more overall good by breaking a rule that it made the most sense on utilitarian grounds to adopt at an earlier point in time, then it should seem clear that a utilitarian must urge you now to break the rule even though accepting the rule was initially justified on good utilitarian grounds. This is what is commonly referred to as the "rule worship problem." The rule worship problem threatens to undermine the rule-utilitarian reply to the punishing the innocent objection even if the rule content problem is successfully overcome. What a defender of punishment seeks to defend, after all, is not the claim that it is morally permissible for the state to *establish* the practice of punishment, but rather the claim that it is morally permissible for the state to actually engage in the practice. But if the rule-utilitarian defense of establishing the practice of punishment without telishment entails that judges should nonetheless engage in punishment plus some form of telishment, then this defense will fail to justify what the defender of punishment seeks to justify.

2.2.2.3.1 TWO CONCEPTIONS OF RULES. Rawls argues that there is a way for the rule-utilitarian to overcome the rule worship problem. His strategy turns on a distinction that he draws between two different conceptions of rules that a rule-utilitarian might employ. The first is what he calls the "summary conception." On this conception, a rule is, in effect, a useful generalization of lessons that have been learned from the past. If, for example, people have found from past experience that breaking promises generally causes more harm than good, this might be used

[30] This problem has been raised by a number of writers, including Duff (1986: 163), Primoratz (1989a: 126ff), and Ten (1987: 70–1).

as the basis for a rule forbidding promise breaking. In the summary sense, the rule gains its authority from the fact that it correctly summarizes the consequences of promise breaking in the past. On this conception of a rule, Rawls concedes, the rule worship objection is successful. If you come across a new situation in which the generalization embodied by the rule is not true, then the fact that following the rule was the right thing to do in previous situations will provide no reason for you not to break the rule in this new situation. If it had been useful in the past to note that promise breaking does more harm than good, for example, but if you then find yourself in a particular situation in which breaking a promise would do more good than harm, there would be no reason for you not to break the rule by breaking your promise.

But Rawls contrasts the summary conception with what he calls the "practice conception" of rules. On this account, rules are pictured as defining a practice. Practices "are set up for various reasons, but one of them is that in many areas of conduct each person's deciding what to do on utilitarian grounds case by case leads to confusion, and the attempt to coordinate behavior by trying to foresee how others will act is bound to fail. As an alternative one realizes that what is required is the establishment of a practice, the specification of a new form of activity; and from this one sees that a practice necessarily involves the abdication of full liberty to act on utilitarian and prudential grounds" (1955: 24).

That is, since the need for the practice arises because of problems caused by people making decisions directly on a utilitarian basis, it follows that the nature of the practice that will satisfy this need will be one that itself prevents people from reasoning in this way.

Rawls then makes two important points about rules so understood. The first is that, on this conception of rules, the "practice is logically prior to particular cases: unless there is the practice the terms referring to actions specified by it lack a sense" (1955: 25). Rawls illustrates this point by appealing to the case of baseball. A person standing on a baseball diamond, for example, can run from third base to home plate and step on the plate. He can do this all by himself. But this, in itself, is not the same thing as scoring a run. An act, that is, can be the act of simply running around the field, and it can be this independently of any rules determining the content of a game. But an act cannot be an act of scoring a run unless there is already a set of rules that determines what counts as scoring a run.

The second point follows from the first: "If one wants to do an action which a certain practice specifies then there is no way to do it except to follow the rules which define it. Therefore, it doesn't make sense for a person to raise the question whether or not a rule of a practice correctly applies to *his* case where the action he contemplates is a form of action defined by a practice" (26). For example, it is a rule of baseball that a

runner is not permitted to step outside the baselines in trying to get from one base to another. If you do this, you are out. And so, on the practice conception of rules, it would make no sense for you to say, "but in this case, wouldn't it be better for me to run outside the lines?" The only way for you to do an act that counts as scoring a run is to do the act in accordance with the rule. This, again, is because the rule is logically prior to the act. The act, understood as a real baseball act, could not exist without it.

This second point has important implications for the punishing the innocent objection:

> [I]f a person is engaged in a practice, and if he is asked why *he* does what *he* does, or if he is asked to defend what he does, then his explanation, or defense, lies in referring the questioner to the practice. He cannot say of *his* action, if it is an action specified by a practice, that he does it rather than some other because he thinks it is best on the whole. ... One doesn't so much justify one's particular action as explain, or show, that it is in accordance with the practice. The reason for this is that it is only against the stage-setting of the practice that one's particular action is described as it is. Only by reference to the practice can one *say* what one is doing. To explain or to defend one's own action, as a particular action, one fits it into the practice which defines it. (27)

And this, in turn, is meant to overcome the rule worship problem: "On the practice conception, if one holds an office defined by a practice then questions regarding one's actions in this office are settled by reference to the rules which define the practice" (28). The result is this: suppose that you are a judge and the rules adopted by the ideal legislators on utilitarian grounds establish the practice of punishment without telishment rather than punishment with telishment. You have before you a particular innocent man whom it would be very useful nonetheless to send to jail. You set him free. You are asked why. Your answer must be that you are a judge and that the acts of a judge can only be assessed from within this framework. And so, according to Rawls, the rule worship problem can be overcome.

2.2.2.3.2 THE PRACTICE WORSHIP OBJECTION. Rawls's argument does seem to establish that the judge, *qua* judge, lacks the legal authority to subject his decisions about who to put in jail directly to a utilitarian standard of evaluation. The judge, in short, should correctly note that his office does not give him the legal authority to punish (telish) innocents in those cases in which this benefits society. He has the legal authority to sentence people, but if he wants to do an act that really counts as an act of sentencing, then he must follow the rules. Otherwise, his act simply can't be an act of sentencing. What makes an act one of sentencing in the first place, after all, is precisely the fact that it takes place within the

legal practice defined by the rules that, in this case, prohibit the deliberate punishment of the innocent. Like the baseball player who wishes to do an act that counts as scoring a run, the judge who wishes to do an act that counts as sentencing someone has no choice but to "stay within the lines" established by the practice that makes the act of sentencing someone possible.

But while Rawls's argument does seem to establish this much, this is not enough to overcome the punishing the innocent objection. After all, we can distinguish between the act of *sentencing* a man to ten years in jail (where it is conceded that this entails that the main is genuinely, or at least genuinely believed to be, guilty) and the act of uttering the words "I sentence you to ten years in jail" in a context in which this will cause an innocent man to be put in jail for ten years (e.g., by causing other relevant legal officials to mistakenly believe that the man has been sentenced and to act on that mistaken belief). Since by definition the latter cannot count as a genuine case of sentencing someone, we must call it something else. Let us call it "telencing." The problem with Rawls's argument, then, is that it establishes only that a rule-utilitarian can show why a judge should not (indeed, cannot) *sentence* an innocent person to jail. It does not, and cannot, establish that a rule-utilitarian can show why a judge should not (or could not) *telence* an innocent person to jail. As Rawls himself puts it, "To explain or to defend one's own action, *as a particular action*, one fits it into the practice which defines it" (1955: 27, emphasis added). So, to justify an act of uttering the words "I sentence you to ten years in jail" *as an act of sentencing*, Rawls seems to be correct in saying that the judge must refer to the rules that define the practice of sentencing, rules that (we are assuming) explicitly forbid the sentencing of an innocent person. But to justify an act of uttering the words "I sentence you to ten years in jail" *as an act of telencing*, one need do no such thing. One need simply justify one's decision to abandon the practice of sentencing in a particular set of circumstances. This is the reason that Rawls's argument ultimately fails to overcome the punishing the innocent objection. The practice conception of rules simply replaces the rule worship problem with the practice worship problem: why should a person go along with a practice in those cases where abandoning it would do more good?

Perhaps an illustration of this point will help to clarify the problem. So, following Rawls, consider a case involving a baseball player. Suppose that you are running from third base to home plate, and you stay within the third-base line even though, as a result, you get tagged out. Suppose that someone then asks you why you didn't run outside the line. Your answer in this case is simply, and correctly, to refer to the practice. You are acting as a baseball player, and in that capacity it is your role to reason from within the rules of the game. It is not your position to question the rules. There is no outside perspective from which to say that it would have been

better for you to step outside the lines. But now, suppose that as you start to run toward home plates, you see a young child in the front of the stands choking on a hot dog. No one around him seems to know what to do, but you know the Heimlich maneuver. So, you run out of the path and save his life. Clearly, *qua* ballplayer, you did something that you were not authorized – not allowed – to do. You were running the bases as a ballplayer and then, suddenly, you did an act that cannot be justified from within the practice of baseball. But it would clearly be absurd to say that, all things considered, you did the wrong thing. Your act can only be described as the act of abandoning the practice, but calling it this does nothing to suggest that you should not have done it. It still seems manifestly clear that you did the right thing. You did what a bad ballplayer but a good human being would do.

The problem for Rawls's argument, then, is this: the judge who finds that he can best promote the public good by putting a particular innocent person in prison is in the same situation as the baseball player who can save the child by running out of bounds. The practice of law can justify to the judge the claim that he must not sentence this innocent person to prison. And it can do so by pointing to the fact that the person is innocent. But the practice of law cannot by itself justify to the judge the claim that he must continue to engage in the practice as defined by the law at this point any more than baseball can justify to the ballplayer the claim that he must continue to play baseball while the child is choking on his food. In particular, the rightness of the legislators' act of instituting the practice of sentencing, which makes it impossible for the judge to *sentence* this innocent person to prison, does nothing to show that he should not *telence* him. And indeed, the very consideration that the rule-utilitarian appealed to in justifying the legislators' act of establishing the practice in the first place now justifies the claim that the judge must abandon the practice in this particular case: namely, that doing so will best promote the public good. A good judge would not telence the person, the rule-utilitarian can say, but he must admit that a good person would. And this concession is all that is needed to undermine the reply to the punishing the innocent objection that is based on the distinction between two conceptions of rules.

A defender of the rule-utilitarian solution might object at this point that I have burdened Rawls's argument with an unfairly high standard of proof. Rawls's job, she might say, is simply to demonstrate that we as a society have a moral right to establish a legal institution in which judges are authorized to punish guilty people and only guilty people. It is not Rawls's responsibility to make sure that judges remain within their legally authorized limits. And if this is so, then Rawls's response to the punishing the innocent objection must be counted a success (assuming that it can overcome the rule content problem) even if the rule worship problem

cannot be overcome. But this response misses the point of the rule worship problem. The problem is not that the rule-utilitarian solution fails to prevent judges from violating their oaths of office. The problem is that it entails that they ought to do so. And it is not enough for a defender of punishment to justify the claim that such offices should be established in the first place. Punishment is not simply a practice by which some human beings take on a certain official role. It is a practice by which they take on a certain role and then act in a certain way as a result. The problem of punishment is to show how it can be morally permissible for people to take on such roles and then to act in such ways. The problem, that is, is not simply to show that we are entitled to establish the practice of punishment. The problem is to show that we are actually to engage in it. This the rule-utilitarian solution cannot do without overcoming the rule worship problem. And, as I have argued, the rule-utilitarian cannot, in fact, overcome this problem. The rule worship problem successfully establishes that the rule-utilitarian solution fails the entailment test (and, for most people, the foundational test as well). For this reason, I conclude that the rule-utilitarian solution to the problem of punishment should be rejected even if my objections to the rule content claim prove unsuccessful.

2.2.3 Other Objections

The rule-utilitarian solution to the problem of punishment was devised to overcome problems with the act-utilitarian solution. The biggest problem with the act-utilitarian solution was that it sometimes justified the punishment of innocent persons. I have argued that, at least on this objection, the rule utilitarian solution ultimately fares no better. Like the act-utilitarian solution, the rule-utilitarian solution must therefore be rejected. Before moving on to consider further versions of the consequentialist solution in general, however, it is worth briefly considering how the rule-utilitarian solution fares with respect to the other objections that I raised against the act-utilitarian solution in Section 2.1. And the answer, I believe, is that it again fares no better.

If I have been correct in maintaining that the rule-utilitarian solution is subject to the punishing the innocent objection, for example, then clearly it will be subject to the not punishing the guilty objection as well. The same considerations that show that it is sometimes useful to punish innocent people, after all, will also show that it is sometimes useful to refrain from punishing guilty people (by finding more useful things for them to do than to be punished, by pretending to punish them, by determining that non-punitive measures such as compulsory victim compensation might produce sufficient deterrence, etc.). A rule-utilitarian, of course, might argue that in the long run, it would be best to enact a system of rules in which judges lack the discretion to refrain from punishing guilty people whenever they

believe that this would produce more good than the alternative. But if my response to Rawls's attempt to fend off the punishing the innocent objection has been successful, this argument, too, will fail, and for the same reasons: utilitarianism would not justify selecting so absolute a rule over one that allowed for some discretion, and even if it did, it would not justify acting in accordance with the rule in cases where breaking it would produce more overall good. In short, the rule-utilitarian solution, like the act-utilitarian solution it is designed to replace, fails to draw the line between those who should be punished and those who should not in the way that the practice of punishment demands that it be drawn. If we acted as rule-utilitarianism would have us act, the result would still look notably different from the practice of punishment.[31]

What about the disproportionate punishment objection and the no excuses objection? Here the problems for the rule-utilitarian solution are, if anything, even worse than the problems for the act-utilitarian solution. If a single act of executing someone for speeding would produce enough deterrence to compensate for the harm done to the speeder himself, for example, then establishing a system of rules in which anyone caught speeding is executed would do so even more effectively. If the act of imposing a relatively modest fine on someone who burns down a $1 million building is enough to produce optimum deterrence for that offense, then adopting a system of rules in which that is the only punishment for that offense will do so as well. And if the act of punishing a provoked attacker more severely than a nonprovoked attacker would best prevent people from attacking others even when provoked, then, once again, setting up a system of rules in which people receive greater punishment rather than less when they have a mitigating excuse will produce more utility than the alternative. In all of these respects, the move from act-based to rule-based considerations will do nothing to blunt the force of the objections. Indeed, since the objections all arise from considerations about what punishment potential offenders will reasonably expect if they commit certain offenses, or commit offenses under certain circumstances, and since the establishment of a clear rule covering such cases will make their expectations clearer than will a mere set of isolated instances of such punishments, the move from act- to rule-utilitarianism will only make the problems for the utilitarian solution to the problem of punishment more severe. These implications, it must be acknowledged, are not, strictly speaking, incompatible with the practice of punishment. They do not, therefore, provide a reason to suspect that the rule-utilitarian solution fails the entailment test. But virtually everyone who wants to justify punishment wants to do so in a manner that avoids such

[31] At least one prominent defender of rule-utilitarianism as a moral theory has acknowledged as much [see Brandt (1995: esp. 82–5)].

implications. These implications will therefore provide further reason to conclude that the rule-utilitarian solution fails a reasonable application of the foundational test. And that is enough to justify rejecting the solution itself.

Finally, it is important to remember the "treating people as a means" objection. I have argued that if we treated people in the way required by rule utilitarianism, the result would look very different from the practice of punishment. Some people who are not permitted to be punished by the practice of punishment would be punished, some who are supposed to be punished by the practice would not be, and others would be punished too much or too little. But let us now suppose that I have been mistaken. Let us suppose that on rule-utilitarian grounds we would adopt and then consistently adhere to a set of rules in which just the right people would be punished and in just the right amount. Even if all of this were so, the rule-utilitarian solution to the problem of punishment would still be subject to one final concern. For it would justify the adoption and following of these rules merely on the grounds that this will be good for society. But in any other context, most people would surely reject the claim that it is permissible to enact and follow a rule that permits the state to intentionally harm some of its citizens simply because enacting and following the rule would best promote social utility. They would not support a rule allowing one group of citizens to be enslaved, for example, or to be used as involuntary experimental subjects, even if there were good reason to believe that the rule would produce better results than any alternative. Those who would reject this reasoning in these other contexts must therefore reject the claim that adopting and adhering to a rule is morally acceptable merely because it produces the best results. And if they reject this claim, then they would still have to reject the rule-utilitarian solution to the problem of punishment even if rule-utilitarianism did entail adopting the practice of punishment.

2.3 OTHER UTILITARIAN VERSIONS

I have argued that the act-utilitarian version of the consequentialist solution is unsuccessful and that the rule-utilitarian alternative ultimately fares no better. But while the act- and rule- formulations are by far the most prominent versions of utilitarianism, they are not the only ones. Before moving on to consider whether or not a nonutilitarian version of consequentialism might provide a more satisfactory solution to the problem of punishment, it is therefore worth briefly considering whether some other version of utilitarianism might suffice.

Consider, for example, what is sometimes referred to as "motive-utilitarianism." On this view, the appropriate subject of moral evaluation is the person's motives: the right thing to do is to have the motives that can

be expected to produce more overall utility than any available alternative. How could the move from act- or rule-utilitarianism to this motive-based version produce a more successful solution to the problem of punishment? A proponent of a motive-utilitarian solution would have to establish two things: that having the motive to ensure that people are punished for breaking the law would produce more overall utility than would having any alternative motive, and that if this motive is the best one to have, then acting from this motive would always be the right thing to do.[32]

But even if motive-utilitarianism proved to be a more attractive moral theory than the more familiar act- and rule- alternatives, I see no reason to believe that it would be a more satisfactory solution to the problem of punishment. The claim that having the motive to ensure that people are punished for breaking the law would produce more overall utility than would having any alternative motive, for example, would be subject to a motive content objection in the same way that the rule-utilitarian solution was subject to a rule content objection: why would the best motive include so absolute a prohibition on punishing innocent people in cases where this would produce more good? Wouldn't an even better motive, from a utilitarian point of view, be one on which one could punish innocents in at least some cases where this would clearly be better, such as cases of vicarious punishment? In addition, even if one was convinced that it would be best to *have* a motive that included an absolute prohibition on punishing innocent people, a motive-utilitarian solution to the problem of punishment would remain open to a "motive worship" objection: why endorse *acting* from the best motive in those cases in which acting otherwise would produce even better results? Even if we thought it best to inculcate in ourselves a motive to never punish innocent people, that is, wouldn't it still be true that in certain cases the right thing to do would be to fight to overcome the inclination that we had (rightly) developed in ourselves so that we could produce even more good by punishing an innocent person? Furthermore, if the disproportionate punishment and no excuses objections are good objections to the act- and rule-utilitarian solutions, won't they be just as effective here? If they succeed in the context of those other solutions, that is, then won't it follow that the best motives to have will include motives to punish some trivial offenses very severely, some severe offenses very trivially, and some offenses more when they are done with an excuse than when done without? And finally,

[32] Some defenders of motive-utilitarianism would reject the second claim. They maintain that in some cases the right act is one that a person with the right motive would not perform. For our purposes, however, this possibility is irrelevant. The problem of punishment is a problem of justifying the set of acts that treats offenders in one way and nonoffenders in another. Motive-utilitarianism is of interest for our purposes only if it can justify this way of behaving, and it can do this only if it justifies not just adoption of the punitive motive but also consistent adherence to it.

even if a motive-utilitarian solution could avoid all of these problems, it would still support punishment merely on the grounds that punishment is a useful way to treat people. Since we would not accept such support as sufficient to render permissible other practices that involve intentionally harming people, it would be wrong to accept such support in the case of punishment, even if motive-utilitarianism were able to provide it.

These reasons for rejecting a motive-utilitarian solution to the problem of punishment, moreover, seem to apply just as readily to any of the other versions of utilitarianism that have been proposed. If one makes dispositions or character traits the ultimate subject of moral evaluation, for example, then the same considerations that were appealed to in rejecting a motive-utilitarian solution would generate a parallel set of reasons for rejecting these further solutions: reasons to deny that it would be best to be disposed never to punish innocent people; reasons to believe that even if it were best to be so disposed, it would be best to sometimes to resist that disposition; reasons to believe that it would be best to be disposed to punish some offenders too harshly, others too leniently, and still others with disregard for their mitigating excuses; and reasons to reject the claim that the mere usefulness of adopting a disposition or character trait could be sufficient to render its adoption morally permissible.

The arguments against the rule-utilitarian solution, in short, provide the basis for arguing against any solution in which a defender of punishment selects a particular aspect of a person's personality or behavior and maintains that the best one of those is the one that maximizes utility.[33] For any version of utilitarianism, therefore, there will be good reason to conclude that the practice that would result from adopting it would not be the practice of punishment; that even if it were, it would not be a suitably proportionate version of punishment; and that even if it were, the fact that adopting it would be useful would be insufficient to render it morally permissible.

2.4 NONUTILITARIAN VERSIONS

I have argued that no version of utilitarianism can provide a satisfactory solution to the problem of punishment. Utilitarianism, however, is not

[33] The same will be true if the utilitarian attempts to choose a particular subset of these or even "the whole complicated set of dispositions, motives and acts over the whole life of an agent." On this last account, "the right life for an agent to try to lead is the life that maximizes expected utility . . . every time an agent makes a choice (where choices can be choices about acts, about motives, about dispositions or anything else) the right choice to make is the one consistent with maximizing the expected utility of her whole life" [Mason (2002: 300, 302)]. And, as with the other positions, there will be good reason to deny that the life such a person would lead would include a commitment to never punishing innocent people and good reason to deny that the mere fact that a life produced the most utility would render it morally permissible.

the only version of consequentialism. Indeed, many people would argue that it is not even the most plausible version. Consequentialism is the view that what is morally right is determined entirely by what best promotes the good. And there are many other views of the good besides utility, and perhaps many more plausible views as well. A consequentialist might, for example, aim to maximize social stability, or social equality, or the extent to which people get what they deserve, or she might offer a pluralistic conception of the good, consisting of a combination of these and other particular goods, weighted or prioritized in one of a number of possible ways. And so, it might be thought, an argument against the utilitarian solution to the problem of punishment could only count, at most, as one very small part of a much larger argument against the consequentialist solution.

If we were looking for the best version of consequentialism in general, we would have to take a long, careful look at the many nonutilitarian conceptions of the good available to the consequentialist. But we are not looking for the best version of consequentialism in general. We are looking only for a solution to the problem of punishment in particular. And for that, we do not have to consider these other versions of consequentialism in detail. The reason for this is simple. The difficulties with the utilitarian solution to the problem of punishment identified in Sections 2.1–2.3 have nothing to do with utilitarianism's account of the good. Instead, they arise from its account of the relationship between the good and the right. And since every version of consequentialism includes this account, every version of the consequentialist solution will be unsuccessful if the utilitarian version fails.

Consider, for example, a version of consequentialism in which the good is cashed out in terms of people getting what they deserve: good people deserve to flourish, bad people deserve to suffer, and one state of affairs is better than another to the extent that a greater proportion of the people in it enjoy the level of well-being that they deserve. On an act- version of this form of consequentialism, an act of punishing a particular offender will be justified if it leads a greater proportion of people to get what they deserve than would any alternative act. And, at least in a good number of cases, we can see why, under such a theory, a particular act of punishing a particular offender might be justified: the offender might deserve to suffer, and punishing him would be likely to bring his suffering about. This version of consequentialism is clearly different from the utilitarian version, and so it might be thought that even if the utilitarian version of the consequentialist solution is unsuccessful, this alternative version might not be.

But this alternative version of the consequentialist solution will fail for precisely the same reasons that the utilitarian version failed. We can construct cases, for example, in which deliberately punishing one

innocent person would do the best job of increasing the proportion of people who get what they deserve. We know that there can be such cases for the same reason that we know that there can be cases in which deliberately punishing one innocent person would do the best job of increasing the overall social utility (e.g., by considering cases where punishing one innocent person would enable the authorities to apprehend and punish five guilty people, or where punishing the innocent child of one offender would deter five other people from wrongfully harming other people's children). We can construct cases in which deliberately refraining from punishing a guilty person would have the same results, and for the same reasons (e.g., cases where not punishing one guilty person would enable the authorities to apprehend and punish five guilty people). In cases such as these, the act- version of this desert-based version of the consequentialist solution will fail for precisely the same reason that the act- version of the utility-based version failed: it will not draw the line between those we may permissibly punish and those we may not in the place that punishment draws it. This will be enough to show that the solution fails the entailment test as well as a reasonable application of the foundational test, and will thus be enough to provide two reasons for rejecting it as a solution to the problem of punishment. In addition, we can construct cases in which the punishment is disproportionate to the offense (either too severe or too trivial); will produce the best results according to the desert-based conception of the good, and, once again, for precisely the same reason that we were able to construct such cases for the utilitarian conception of the good. And so, the act- version of this desert-grounded version of the consequentialist solution will run into precisely the same problems in passing the foundational test as the utility-based version. And, of course, in doing all of this, this further version of the consequentialist solution would still harm people simply as a means of producing a greater amount of good in precisely the same way that the utilitarian version did.

The lesson to be drawn from this example of a nonutilitarian version of the consequentialist solution can be generalized in two ways. First, it can be applied, and with the same force, to the act- version of any version of consequentialism. For any good (or set of goods) that consequentialism holds out as the good to be maximized, if punishing an offender usually promotes that good, then there will always be cases where deliberately punishing an innocent person will also promote that good (again, e.g., by increasing the number of guilty people who get punished) and cases where refraining from punishing a guilty person will promote it (for the same reasons). The widely discussed law and economics approach to punishment, for example, construes punishment as a means of maximizing social wealth, rather than utility, by imposing an expected cost on potential offenders to offset the costs that society pays for their offenses

[see, e.g., Avio (1993: 250–62)]. But the same reasons for holding that punishing the innocent will sometimes produce more utility apply just as forcefully when the goal is maximizing social wealth. A number of writers have attempted to justify punishment by appealing to the claim that it minimizes the private pursuit of vengeance [e.g., Perkins (1970)], but if this is the good that the consequentialist seeks to use in attempting to justify punishment, it will again justify punishing innocent people, this time in cases where not doing so would lead to more acts of private vengeance (e.g., because most people think that the innocent person is guilty and would therefore seek vengeance against both him and the government officials who let him go).[34] Punishing the innocent is a problem for the act-utilitarian solution, in short, not because the act-utilitarian has an objectionable account of the good, but because the position makes the permissibility of each act of punishment turn on the expected consequences of that act. Since this is a feature of every act-consequentialist position, it follows that every act- version of the consequentialist solution to the problem of punishment will be subject to the punishing the innocent objection. In addition, on any account of the good to be promoted, there will be cases in which disproportionately great or disproportionately trivial punishments will produce the greatest overall good, again for the same reasons. If imposing a great loss of the good on one offender, whatever the good may be, can be used to prevent an even greater total loss of that good by many others, then the act- version of the consequentialist position will endorse great punishment even for minor offenses, regardless of the conception of the good that it appeals to. And whatever the conception of the good it appeals to, of course, a justification of punishment based on the claim that punishment would maximize the attainment of that good will also involve harming people merely to produce a greater amount of that good. And so, if the act-utilitarian solution to the problem of punishment is unacceptable, then so is the act-version of any consequentialist solution.[35]

The second way in which the lesson drawn here can be generalized is that it can be applied to non-act- versions of any version of the consequentialist solution. The problems with the rule-utilitarian solution, for

[34] The same is true of the related proposal that punishment is justified as a means of maximizing conflict resolution [Cragg (1992) can be understood as defending punishment in this way; for a critique of Cragg, see Brunk (1996)].

[35] It might be thought that there is a conception of the good that would avoid these problems: one that consists of, or at least includes, the view that deliberately punishing innocent people is itself a bad thing. But even on this version of the good, there will be cases in which deliberately punishing one innocent person would produce more of this good. For example, there could be cases in which corrupt judges deliberately punish innocent people, and the best way to deter them is to frame and punish an innocent judge.

example, were once again completely independent of the utilitarian conception of the good. The rule content and rule worship objections did not arise because there was a problem with the claim that the best states of affairs are those that contain the most social happiness. Rather, it arose because there were problems in moving from that claim to the conclusion that we should adopt a rule that includes an absolute prohibition on deliberately punishing innocent people, and problems in moving from that conclusion to the further conclusion that if we should adopt such absolutist rules, then we should also adhere to them even in cases where breaking them would do more good. These problems arose not because there was a problem with the utilitarian account of the good, but because there were individual cases in which punishing an innocent person would produce more of that good. Because such cases existed, there were reasons for the utilitarian to prefer rules permitting some punishment of the innocent to rules permitting none at all, and to prefer sometimes breaking the rules to never breaking them. The rule-utilitarian solution, in short, was ultimately unacceptable because there were individual cases in which deliberately punishing the innocent would produce more utility. And since, as I have argued, in any version of consequentialism there will be individual cases in which deliberately punishing an innocent person would produce more of the good, it follows that the rule- version of the consequentialist solution will fail for any version of consequentialism.[36]

[36] Lippke, for example, attempts to defend the permissibility of punishment by appealing to a consequentialist framework on which the good to be promoted is the good of preserving "a system of equal rights," which in turn generates various goods such as "various kinds of liberties and enjoyments." He then suggests that punishing people for violating the rights of others might prove "indispensable" to preserving such a system and that, as a result, we would "have sufficient reason to endorse a scheme of legal punishment as a way of securing for ourselves the benefits of a system of equal rights" (2001: 85). But even if we agree that we should maximize the preservation of a system of equal rights, the rule content and rule worship objections will undermine the claim that this could be used to justify punishment: a rule occasionally permitting punishment of the innocent might produce even better results by this measure (e.g., a rule that would permit a judge to punish one innocent person as a means of deterring others from violating equal rights, such as framing and convicting one innocent police officer to deter others), and even if we did have sufficient reason to "endorse" an absolute rule against such punishment, the goal of preserving such a system might in some cases warrant breaking the rule.

The same problem undermines the attempt of Braithwaite and Pettit to justify punishment on consequentialist grounds by appealing to the good of maximizing people's ability to enjoy what they call "dominion," by which they mean a form of control over one's life that involves "being free from the interference of others" such that one "has that control in virtue of the recognition of others and the protection of the law" (1990: 60). Braithwaite and Pettit argue that the goal of maximizing dominion so understood cannot be met if government officials aim directly in each instance to maximize it, but their "rule-dominion" solution is ultimately unsuccessful for the same general reasons that undermine the rule-utilitarian solution. They argue that "if the

Finally, since the objections that I raised against other possible versions of the utilitarian solution in Section 2.3 were independent of questions about the utilitarian conception of the good, it will follow that if, for example, the motive-utilitarian solution is unsuccessful, then so is any motive- version of the consequentialist solution; if the disposition-utilitarian solution is unsuccessful, then so is any disposition- version of the consequentialist solution; and so on.

The results of this chapter can therefore be summarized as follows: the act-utilitarian solution to the problem of punishment is unacceptable, and if that solution is unacceptable, then so is every version of the consequentialist solution. Every version of the consequentialist solution is unacceptable because every version of consequentialism is entirely forward-looking, and on any account of the good to be aimed at in the future, there will be cases in which more future good will be produced by punishing some innocent people and not punishing some guilty people, in which more good will be produced by punishing some guilty people too much and others too little, and in which people will be harmed merely to produce a greater overall amount of good. In short, no version of the consequentialist solution passes the entailment test, and for most believers in punishment, none will pass the foundational test either. And thus, no version of consequentialism can provide an adequate solution to the problem of punishment.

criminal justice authorities are bent on promoting dominion, then their responses must not *always* be determined by direct consideration of that target. If they pursue the promotion of dominion free of *any* constraints, then they will not promote it" (74). But while this makes plausible the claim that one who seeks to promote dominion should not be an "act-dominionist," it does nothing to show that dominion will best be promoted by adopting absolute constraints on punishing an innocent person to deter others from infringing on the dominion of others (again, e.g., by framing one innocent police officer in order to deter many others), nor does it show that people should always follow such rules, even if they should adopt them. For further objections to Braithwaite and Pettit's argument, see Matravers (2000: 26–9).

3

The Retributivist Solution

3.0 OVERVIEW

The consequentialist solution to the problem of punishment is essentially forward-looking. It attempts to justify punishing offenders in the present by appealing to the beneficial effects that this will bring about in the future. I argued in Chapter 2 that this solution is unsuccessful. The retributivist solution, by contrast, is essentially backward-looking. It claims that committing an offense in the past is sufficient to justify punishment now, whether or not this will produce any beneficial consequences in the future. Some people accept the retributivist justification of punishment simply because they are repelled by the consequentialist alternative. Primoratz, for example, argues that it is reasonable to "set out from the assumption that the institution of punishment is not unjustifiable in principle" and then to justify the retributivist position by identifying the unacceptable "implications of the competing justifications of punishment" (1989a: 148). Leslie T. Wilkins justifies his endorsement of the retributivist conclusion of the *Report of the Committee for the Study of Incarceration* by explaining that "I cannot accept it as a declaration of a desirable policy – it is merely less unacceptable than any others which can be considered at this time" [quoted in Von Hirsch (1976: 178)]. And Moore has expressed sympathy for what he calls the "reluctant retributivist" who becomes "a retributivist by default" because he finds "decisive objections" to all the other attempts to justify punishment (1987: 97). On this account, one is led to accept the retributivist justification of punishment simply because all the other justifications of punishment are even worse.

The reluctant retributivist begins with the assumption that punishment is morally permissible. Reluctant retributivism, therefore, cannot constitute a solution to the problem of punishment. It can count only as an evasion of it. The same is true of what Hill refers to as "deep retributivism," the view

on which the claim that offenders should be made to suffer for their offenses is "a fundamental principle in need of no further justification" (1999: 409).[1] The deep retributivist does not offer a solution to the problem of punishment but instead insists that it is self-evident that there is no such problem in the first place. This response is unsatisfactory because many people, including many who believe in punishment, do not find retributivism to be intuitively self-evident; indeed, many of them find it counterintuitive.[2]

But reluctant retributivism and deep retributivism are not the only strategies available to the proponent of the retributivist solution.[3] There are two other approaches that can be, and have been, used in attempting to justify, and not merely to endorse, the retributivist position. The first approach starts with particular judgments about particular cases and moves to a more general principle of retributive justice. The second starts with an even more general moral principle, such as a principle of rights or of fairness, and derives from it a more specific principle of retributive justice. I will consider the first approach in Section 2, where I will examine the claim that the retributivist solution can be extracted from particular judgments we are likely to make about what particular people deserve. I will consider the two most prominent versions of the second approach in the two sections that follow, where I will examine the claim that the retributivist solution can be derived from a more general theory of rights (Section 3.2) or of fairness (Section 3.3). Finally, I will consider some less familiar but potentially powerful further versions of the second approach, by which some recent writers have attempted to ground retributivism in general considerations about trust (Section 3.4.1) and about debt (Section 3.4.2), and will conclude by saying something about

[1] As Hill notes, Kant has frequently been understood to hold this view, although Hill himself rejects this interpretation. See also Hill (1997: 194ff.).

[2] In addition, as Golash has pointed out (1994: 73), even if one does have the retributivist intuition that the wrongness of the offender's act suffices to justify the offender's punishment, one will also have the intuition that punishment involves treating the offender in ways that are typically impermissible. And this should be enough to render unsatisfactory the appeal to intuition as a foundation for the retributivist solution.

[3] In a frequently cited essay, "Varieties of Retribution," John Cottingham identified nine theories of punishment that have sometimes gone under that name (1979), and in a more recent essay, Nigel Walker distinguishes even more (1999). Up to the early 1970s, retributivism was a relatively minor strand in the literature on punishment, which was dominated by the consequentialist position. That situation began to change with the publication in 1976 of two books: *Doing Justice*, by von Hirsch, and *Fair and Certain Punishment*, the report of the Twentieth Century Fund Task Force on Criminal Sentencing. The tide has changed so significantly since then that by 1990 one student of the subject could plausibly write, in a survey on recent work in the field, that "today, the theory of punishment is largely retributive theory" [Davis (1990a: 220); similarly, see Ellis (1995: 225)].

the relationship between retributivism and revenge (Section 3.4.3). I will argue that none of these attempts to defend the retributivist position is satisfactory and that neither, therefore, is the retributivist solution to the problem of punishment.

3.1 DESERT-BASED RETRIBUTIVISM

The most straightforward version of the retributivist solution to the problem of punishment is based on the concept of desert: punishing people for breaking the law is morally permissible because such people deserve to be punished. The most important defenses of this desert-based form of retributivism are those of Michael S. Moore and Stephen Kershnar.[4] I will therefore organize this section around a discussion of their arguments, supplementing it with work by others who have defended a similar view. Moore and Kershnar, like many other retributivists, attempt to defend two distinct claims: that we have a right to punish offenders and that we have a duty to do so. Since only the first claim is needed to provide a satisfactory solution to the problem of punishment, I will focus here entirely on their defense of it.

3.1.1 The Argument from Cases

The desert-based defense of the claim that it is morally permissible for the state to punish people for breaking the law can best be understood as an argument from particular cases. The proponent of the theory, that is, starts by focusing on examples to which he assumes we will all have a certain intuitive response. He then argues that the theory that does the best job of accounting for these judgments involves a principle of desert, a principle that, in turn, can then be used to justify the permissibility of punishment.[5] If this is true, then punishment turns out to be morally permissible because the claim that it is morally permissible is needed to account for the correctness of several more specific judgments that are assumed to be true. Both Moore and Kershnar proceed by trying to establish that a desert-based form of retributivism "best accounts for" [Moore (1987: 98)] or provides the "best explanation of" [Kershnar (2001: 41, 74–5)] these moral judgments [see also Moore (1993: 24)].

The judgments that the desert-based retributivist appeals to in making this argument typically involve our reactions to horrendous atrocities committed by unrepentant offenders. Moore, for example, cites several

[4] See, e.g., Moore (1987, 1993) and Kershnar (2000, 2001). Desert-based retributivism is also defended by Hawkins (1944), Lewis (1949) Mundle [(1954), though later rejected by Mundle (1968)], Kleinig (1973: esp. chap. 4), and Primoratz (1989a: chap. 7).

[5] See, e.g., Moore (1987: 98, 120) and Kershnar (2000: 101, 114).

such examples from the writings of newspaper columnist Mike Royko, involving (among other things) a man who "raped and murdered [a stranded motorist] and drowned her three small children, then said that he hadn't been 'losing any sleep' over his crimes," an armed robbery at the end of which "[f]or no reason, almost as an afterthought, one of the men shot [and killed] the grocer," and a kidnapper who "kept [his victim] in the trunk [of a car], like an ant in a jar, until he got tired of the game. Then he killed her" (1987: 98–9). Murphy, in a related way, appeals to the reactions that audiences have to the most vicious villains in movies (1990b: 64–5). And virtually everyone who defends retributivism on desert-based grounds makes use at some point of the case of the (usually unrepentant) Nazi war criminal [e.g., Kleinig (1973: 67); Primoratz (1989a: 149)].

Brutal murderers, movie villains, and Nazis, of course, represent the extreme end of the spectrum of antisocial behavior. But if the desert-based retributivist can succeed in using them to elicit intuitions favorable to his position, he seems to be in a position to use them as a way of moving us toward similar judgments in less extreme cases. As Primoratz puts it:

> If we accept this [desert-based retributivist] claim with regard to the crimes committed in Auschwitz and Buchenwald, why not accept it with regard to crimes against humanity of lesser magnitude? And if we accept the claim with regard to the latter as well, why not with regard to murder of a single human being? And with regard to other crimes, less serious than murder? Or, if we are not willing to go all the way with this demand that justice be done and the criminal paid back in full, where, precisely, shall we draw the line? (1989a: 149)

The question, then, is whether the desert-based retributivist can succeed in arguing from the clear and extreme cases about which virtually everyone seems to agree to the conclusion that punishment itself is morally permissible.

I will argue in the subsections that follow that the answer to this question is no. First, however, it is necessary to be clear about just what it is that everyone is supposed to agree on the first place. The argument from cases, that is, attempts to defend the desert-based retributivist solution on the grounds that "it best accounts for those of our more particular judgments that we also believe to be true" [Moore (1987: 98)], and we must begin by clarifying what those particular judgments by retributivists like Moore and Kershnar are supposed to be.

One possibility is that they are judgments to the effect that it is morally permissible (indeed, obligatory) to punish these particular people. At one point, for example, Moore refers to the judgments that he is counting on the reader to share by saying that "[m]ost people react to such atrocities

with an intuitive judgment that punishment is warranted" (1987: 99). At another point, referring to the nobleman in Dostoevsky's *Brothers Karamazov* who has his dogs tear a young child to pieces in front of the child's mother, Moore appeals to the intuitive judgment that the nobleman "should" be punished (1993: 25, 29). Kershnar, too, at times characterizes these judgments by saying that it "intuitively seems that [a person in a particular example] should be punished by the state" [(2000: 101, 104); see also (2001: 78–80)].[6] On this version of the argument from cases, the desert-based retributivist is counting on you to believe that punishment is morally justified in at least some cases, and is then attempting to generalize from such cases as a means of justifying the practice of punishment in general.

If this is what the judgments used to ground the argument from cases ultimately amount to, however, then there is no way for the argument to provide a solution to the problem of punishment. Moore, for example, is presumably right to expect most people to believe that the offenders should be punished in the cases he refers to. If people did not have this belief, after all, there would be no problem of punishment in the first place. The problem of punishment, that is, arises not because people do not have this belief, but precisely because they do have it and are then led to wonder if the belief can be justified. And so, appealing to the fact that people already believe that punishment is justified in these cases cannot serve as a basis for solving the problem. It can serve only as a reminder that the problem exists.[7]

Moore attempts to respond to this kind of concern about this formulation of the argument from cases by emphasizing that the judgments in

[6] Similarly, on Garcia's analysis of negative desert, to say that an offender deserves to be punished by the state amounts to saying that the state would not violate his rights by punishing him (1986: 223; 1990: 153). Relatedly, Mackie suggests that the belief that it is morally permissible to punish the guilty is "very widely, perhaps universally, felt to have … an immediate appeal and underived authority" (1982: 4). Mundle, too, seems to endorse a form of desert-based retributivism while conceding that it cannot be defended beyond appealing to the fact that many people seem to believe it (1954: 227–28).

[7] Kleinig can perhaps be understood as attempting to offer a non-question-begging argument from such intuitions. In answer to the question of whether the unrepentant Nazi war criminal would be justified in complaining if he was later made to suffer for his atrocities, Kleinig writes: "I believe not, precisely because it would be just that he should so suffer. To think otherwise would be tantamount to believing that it was all right to do what you like so long as you could get away with it, or at least get away with it for long enough and in such a way as to deprive punishment of any useful consequences" (1973: 67). This, in effect, is to offer a version of the logical entailment argument examined in Section 1.2.1: to believe that what the Nazi does is not "all right" entails that it must be just that he suffer for it. But, as we saw in Section 1.2.1, such entailment arguments are unsuccessful. One can believe that acts do not merit punishment without believing that they are "acceptable if, for example, one believes that they merit censure or compulsory restitution, that the state may forcibly prevent them from being performed, and so on.

question are not of the form "the nobleman should be punished because and only because he deserves it" but rather of the form "the nobleman should be punished, period" (1993: 29). If the argument relied on our assenting to judgments of the first sort, he concedes, the argument would be objectionably circular, but this is not so given that it relies only on our accepting judgments of the second sort. On this second understanding, the argument can still "possess a non-trivial justificatory force for sustaining the retributive principle," according to Moore, because it allows us to abstract away from the particular cases and infer what they all have in common: that the offenders deserve to be punished, where "'desert' means culpable wrongdoing" (29). The problem with this response, however, is that it still enables the argument from cases to produce nothing more than what we already know: that we believe that culpable wrongdoers should be punished. Moore is correct to insist that the argument's justificatory force is nontrivial in the sense that what all the cases had in common might have been something else. But this is not enough to serve as a basis for solving the problem of punishment because the problem itself arises precisely from the fact that we already know that we draw this particular line and treat it as morally relevant: what we want to know is why it is permissible for us to do so.

Since this reason for rejecting the first version of the argument from cases may not be immediately apparent, an analogy may be useful. So, consider the question of whether or not it is morally permissible for one person to kill another in self-defense. Virtually everyone agrees that, at least under certain conditions (the threat to the person is very serious, there is no nonlethal alternative that will eliminate the threat, etc.), such killings are morally permissible. But, at least on the face of it, the belief that killing a person in self-defense is morally permissible poses a problem. Most people, after all, agree that killing people in general is morally impermissible. The question then naturally arises: why is it permissible to kill people in these circumstances when it is impermissible to kill them in most others? There are a number of answers that might be given to this question. One could argue, for example, that a person who initiates an attack on another thereby forfeits his right to life; that if one or the other of the two inevitably will be harmed, then fairness dictates that the harm should be incurred by the one who is responsible for the attack; that killing in self-defense involves only foreseeable harm and not intentional harm; and so on. Any of these answers might in the end prove sufficient. But clearly, it would not be sufficient to attempt to answer this question by simply describing cases in which one person is attacked by another, noting that most people will believe that killing the attacker in such cases is morally justified, and then inferring from this a general principle that killing in self-defense is morally justified. In the context of the problem of self-defense, it would be clear that this would not solve the

problem; it would simply highlight it by demonstrating that most people do, in fact, treat the line between killing in self-defense and killing in other contexts as morally relevant. Yet, this is precisely what the defense of desert-based retributivism is reduced to if the argument from cases is understood as building on the fact that most people already believe that punishment is justified in the kinds of cases that Moore and others appeal to. And so, it should be clear that the argument from cases, so understood, cannot provide a satisfactory solution to the problem of punishment.[8]

A second way of understanding the argument from cases is suggested by another way that Moore characterizes our presumed response to the cases he cites: "most of us . . . feel some *inclination*, no matter how tentative, to punish [in such cases]. That is the particular judgment I wish to examine" (1987: 99, emphasis added). On this second construal, the argument from cases begins not with the question-begging claim that punishment is morally permissible, but rather with an inclination to punish in such cases. Since this inclination to punish is likely to be as widespread as the corresponding belief about the permissibility of punishment, and since the inclination does not beg the question in favor of the belief, this second construal of the argument from cases renders it immune to the objection that undermines the first. But this second construal is unsuccessful nonetheless. For not only does having an inclination to punish not beg the question in favor of a belief that punishment is permissible, it does not even provide support for that belief. There are, after all, many inclinations that many people naturally have – inclinations to feel envy, jealousy, lust, vindictiveness – that they do not consider permissible to act on. And so, if all the argument from cases elicits from us is an urge to punish such people, this urge will not take us far in attempting to solve the problem of punishment.

What the argument from cases requires, therefore, is that the cases it appeals to elicit from us a judgment that is a genuine belief and not simply a feeling – but one that goes beyond the belief that punishment is morally justified. And there is, in fact, one final way of construing the argument to meet this requirement. At another point in Moore's article, he characterizes the judgments he wishes to argue from in yet another way: "I suspect that almost everyone at least has a tendency . . . to judge culpable wrongdoers as *deserving of* punishment" (1987: 98).

On one possible interpretation of this claim, of course, to say that a person deserves to be punished is simply another way to say that she should be punished. And on this interpretation, it should be clear, this third construal of the argument from cases will simply collapse into the

[8] It may also be worth noting that even if my objection to this first version of the argument from cases is rejected, this version will still fall prey to the first two objections that I raise against desert-based retributivism in Sections 3.1.2 and 3.1.3.

first and will be unacceptable for the same reason. But saying that a person deserves to be punished can mean something importantly different from this, and on the strongest reading of the argument from cases, it does. On this alternative reading, saying that a person deserves to be treated in a certain way does not *mean* the same thing as saying that he should be treated in that way, but saying that he deserves to be treated a certain way nonetheless provides *support* for the claim that he should be treated that way. It provides support for this claim, on this account, because in judging the overall value of a particular state of affairs, it maintains that we should not simply consider how much happiness or unhappiness each person in it has, but should also consider the positive and negative desert of those who have it. We should think, that is, not that happiness itself is good but that deserved happiness is good, not that unhappiness is bad but that undeserved unhappiness is bad. To say that a person deserves to be punished, on this account, is not simply to say that she should be punished. Rather, it is to say that the world will be intrinsically a better place by one morally relevant measure if she is punished than it will be if she is not. As Mundle puts it, "the state of affairs in which an offender is punished is less evil than that in which he goes unpunished" (1954: 74). Kershnar, too, often puts desert claims in terms of good and bad states of affairs (e.g., 2000: 98–9; 2001: 2–5). And Moore himself, at certain points, is quite explicit in saying this: "what *is* distinctively retributivist is the view that the guilty receiving their just deserts is an intrinsic good," where "[a] state of affairs, act or practice is intrinsically good if its goodness does not depend on some further effect that the state of affairs, etc., produces" (1993: 19, 22).[9]

The belief that this is so in the sorts of cases that retributivists like Moore and Kershnar appeal to is presumably widespread, and if we take

[9] See also Gendin (1970: 7). This view can, in turn, be understood as part of the more general view that the world is a morally better place when the good thrive and the wicked suffer. See, e.g., Pojman (1999: 93): "It is intuitively obvious that the appropriate distribution of happiness and unhappiness should be according to virtue and vice." In addition, although Nozick sometimes presents his defense of retributive punishment in the language of the moral education theorist (e.g., 1981: 370, 372), his position, too, ultimately comes down to the claim that it "is in itself good" when punishment has negative consequences for a wrongdoer (377; see also 379). Nozick views what is good about the resulting state of affairs in terms of its being good that correct values ultimately have an impact on the wrongdoer rather than in terms of its being good that the wrongdoer suffers (e.g., 374, 375–6), but on either account the fundamental position is the same: punishing a wrongdoer is good because there is something intrinsically better about the state of affairs in which a wrongdoer is punished than about one in which he is not. The objections developed in this section against desert-based retributivism, therefore, will tell against Nozick's version of retributivism as well, although he does not characterize his position as being grounded in desert [for a useful, though brief, set of further worries about Nozick's discussion, see also Walker (1995)].

it to be so, it can then provide the basis for a non-question-begging version of the argument from cases. When we consider the cases cited by Moore, Kershnar, and others, that is, we will agree that in such cases, the state of affairs in which the offender is punished for his offense is better than the state of affairs in which he is not punished, and we will use this agreement as the basis for concluding that the offender should be punished (or, at least, for concluding that it would be permissible to punish him) in such cases. And in doing all of this, it is important to emphasize, we will be reasoning in a manner that is retributivist rather than consequentialist. The claim we will be led to endorse, that is, is not the claim that punishing a particular offender will have positive consequences in the future that will, in turn, render the act morally justified in the present. Rather, it is the claim that punishing a particular offender is justified regardless of whether the act produces any further positive consequences in the future simply because his being punished right now is in itself better than his not being punished. Punishing the offender who deserves it will, on this account, be constitutive of the world's being better, rather than merely serve as a cause of some further effect that will make it better. As Moore puts it, "Even if punishing the guilty were without any further effect, it would be a good state to seek to bring about, on this intrinsic goodness view of punishing the guilty" (1993: 20). Some philosophers, it is worth noting, have challenged this judgment, wondering how it could ever be good in itself that someone suffer.[10] While I have some sympathy for this point of view, I will concede here, at least for the sake of the argument, that the desert-based judgment being used to ground the argument from cases is correct. The question, then, is whether it is sufficient to ground a solution to the problem of punishment. I will argue that it is not.

3.1.2 The Not Punishing the Guilty Objection

Punishment, as we have seen, treats the line between those who break the law and those who do not as morally relevant. A successful solution to the problem of punishment must therefore explain why it is morally permissible to draw such a line in just this place. A proposed solution that proves unable to do this will thereby fail the entailment test that any solution to the problem of punishment must pass and will, in addition, strike most people as failing the foundational test as well. I will argue in

[10] E.g., Narveson (1974: 191–3), Hampton (1984: 141), and Hill (1999: 425). See also Blanshard (1968: 75), who argues that the good deserve honor and respect, while the bad deserve contempt and disapproval (but not suffering), and Shafer-Landau (2000), who defends what he calls the thesis of "nihilism about moral desert," in which there is simply no fact of the matter about how much suffering a particular offender deserves.

this section and in the section that follows that the desert-based version of the retributivist solution is unable to draw the line where the practice of punishment draws it for two reasons.

The first reason is that there are people who break a just and reasonable law but who do not morally deserve to suffer. Such cases give rise to the not punishing the guilty objection. One kind of case that gives rise to this problem involves people who do acts that are illegal but not immoral. This may be because the act that the law forbids is never in itself an immoral act, or it may be because although the act is generally immoral, it is not so in a particular case. A person who uses her car to drive a friend to the emergency room even though her car has not passed a required emissions test, for example, is clearly violating the law but is just as clearly behaving morally. A person who steals a car to drive a friend to the emergency room clearly breaks the law, but at least in the case in which the friend will otherwise die and the owner of the car will not be significantly harmed by the theft (the owner, for example, has two other cars sitting in his driveway), it again seems clear that the act is not morally wrong. Reasonable people may disagree about the merits of a particular case, but it is difficult to avoid the conclusion that, on any reasonable view, there are many acts that are not immoral even though they violate morally justified laws. In order to be reasonable, after all, laws must be sufficiently clear and determinate to be understood by those whose behavior they are designed to govern. And such laws, by their very nature, are unable to circumscribe all and only those instances in which an act of a certain type would be morally objectionable. A law that says "do not steal cars," for example, is reasonable in a way that one that says "do not steal cars unless doing so would produce a moral good of sufficient value to outweigh the prima facie wrongness of stealing" is not. Unavoidably, therefore, there will be cases in which an act is not immoral even though it violates a just and reasonable law.[11] But if an act is not immoral, then the person who performs it does not deserve to suffer for it. And if the person does not deserve to suffer for it, then the desert-based retributivist position cannot justify the claim that the state has the moral right to punish him for doing it. Such cases, therefore, establish that desert-based retributivism cannot draw the line between those who may be punished and those who may not in the place that the defender of punishment must draw it.

There is a second kind of case that can also be used to support the not punishing the guilty objection, but before turning to it, let us briefly consider two responses that might be raised against my appeal to the first kind. One response is based on the claim that there is a moral obligation

[11] For a compelling discussion of this feature of laws, see Alexander and Sherwin (2001: esp. chap. 4).

to obey the law, at least on the assumption that the law itself is just and reasonable. If there is a moral obligation to obey the law, then every single act of breaking the law will involve a violation of a moral obligation. If this is so, then there are no acts that are both justifiably illegal and not immoral. And if this is so, then every offender will deserve to suffer for her offense. The problem with this response to the not punishing the guilty objection can best be put in the form of a dilemma: if there is a moral obligation to obey the law, then it is either an obligation that can be overridden by other moral considerations or it is not. If the obligation can be overridden, then there will still be cases in which an act is illegal but not immoral, namely, those cases with further moral considerations that override the obligation to obey the law. If the obligation cannot be overridden, then the desert-based retributivist can show that everyone who breaks the law deserves to be punished, but he will also commit himself to implications that virtually everyone will find unacceptable. It seems legitimate for the state to require periodic emissions tests for cars, for example, and reasonable to suppose that this requirement would become unworkable if it contained the provision that a car could be driven without such tests whenever the driver had a morally good enough reason to do so. And so, on the view that the obligation to obey the law cannot be overridden, it would be morally impermissible to drive your friend to the emergency room if your car had not yet passed its emissions test. And this seems plainly to be an unacceptable result.

The second response that might be given to the first case I appealed to in defending the not punishing the guilty objection appeals instead to the claim that all legal prohibitions should follow moral prohibitions. There should be laws against rape, murder, theft, and so on, on this account, since such behaviors are immoral; but if, for example, there is nothing immoral about smoking marijuana, then there should be no laws against it. If the only laws that are just and reasonable forbid acts that are independently immoral, then it will again follow that every justifiably illegal act is also immoral. The problem with this second response is that there are too many laws that seem just and reasonable even though they forbid behaviors that are not independently immoral. It is not immoral to drive (safely) without a license, for example, or to put an addition on your house without first obtaining a building permit, but most people will agree that it is just and reasonable to have laws that prohibit such acts nonetheless. Even if we think that people should generally obey the law, therefore, and that the law should generally allow behaviors that are not independently immoral, there are many acts that are not immoral even though they violate just and reasonable laws. In these cases, a defender of punishment must be prepared to explain why the state has the right to punish the offender, and the desert-based retributivist will be unable to do so.

I have argued that some acts are not immoral, even though they violate just and reasonable laws, and that this poses a difficulty for the desert-based retributivist position. But it is important to recognize that even if I have been mistaken about this – even if every single illegal act also turns out to be immoral – it is still not true that everyone who breaks the law deserves to suffer. This is because whether or not a person deserves to suffer is ultimately a function of whether she acts from a good or bad motive rather than whether her acts are objectively right or wrong. A good person may do an objectively immoral act because she believes it to be the right thing to do, for example, and although in such cases we should certainly condemn her behavior, it does not follow that we should think of her as a bad person or as deserving of punishment. Indeed, at least in those cases where we do not think the person blameworthy for the mistaken belief she acts on, it seems clear that she does not deserve to suffer, although her act does deserve to be repudiated.[12]

Suppose, to start with a simple case, that Larry has recently moved into a new neighborhood where all the houses look alike. The only way he recognizes his own house is by the large birdbath on the front lawn. One day while he is at work, his wife meets their next-door neighbor and gives her the birdbath as a token of friendship. When Larry gets home late at night, he reasonably believes that the house with the birdbath in front is his, and as he walks up the sidewalk, he sees a young woman trying to get into the house. When he demands to know what she is doing, she responds that she does not speak English. Having heard that several houses in the area had recently been broken into by a woman claiming to speak no English, Larry prevents her from going into the house. As it turns out, however, the woman is not a burglar and is simply trying to get into her own home. Objectively, therefore, Larry's act is clearly wrong. But it seems extremely unlikely that anyone would respond to the situation by saying that Larry deserves to suffer for having done this

[12] Kershnar has defended what he calls the "act theory of deserved punishment," in which it is the performance of a certain act, rather than having a character of a certain sort, that is necessary and sufficient for deserving punishment (1997b; 2001: 17–24). If this theory is correct, then this second version of the not punishing the guilty objection must be rejected. But the cases that Kershnar appeals to in considering the merits of the character-based account relative to the act-based account are importantly flawed. In particular, in arguing that a bad act is sufficient to deserve punishment even when done by a person with good character, Kershnar appeals to cases in which a good person uncharacteristically gives in to a bad motive [e.g., "a virtuous person with a lifetime of good works who temporarily succumbed to greed" (1997b: 510); see also (2001: 29ff.)]. We can agree that in this case the person deserves to suffer for his greedy action while still maintaining that good people who do bad acts out of *good* motives do not. And that is all that is needed to sustain the second version of the undeserving offender objection [the same problem also undermines the cases that Sendor appeals to in arguing for the claim that character is not relevant to desert (1996: 129–32)].

objectively wrong act. And the reason for this is simple: although objectively the woman had every right to go into the house she was trying to enter, Larry sincerely and nonnegligently believed that she did not. Since he acted in good faith on the mistaken belief that she had no right to enter the house, his act does not cast him in a bad light, so we do not think that he deserves to suffer for what he did.

But now suppose that Moe prevents a pregnant woman from entering an abortion clinic or a female employee from entering a nuclear weapons plant. If we assume that these women have the right to enter the buildings in question (at least in one or the other of these cases), then these acts of blocking their access are also objectively wrong. But if Moe's act is based on having thought a great deal about the relevant issues and having determined that shutting down the clinic or plant is necessary to prevent the deaths of many innocent persons, then he is as morally blameless for his act of obstruction as Larry is for his. Moe, like Larry, actively prevents a woman from entering a building because he sincerely believes that she has no right to enter. In the case of Larry, it seems clear that since his motivations are unobjectionable, he does not deserve to suffer for his behavior. The same, then, must be said of Moe. But on the assumption that laws forbidding such obstruction are just and reasonable,[13] these, too, will be cases in which the practice of punishment dictates that the state should have the right to punish and the desert-based retributivist solution will be unable to account for this.

Indeed, even in more extreme cases of objectively immoral behavior, I find it difficult to accept that a person who acts out of good intentions deserves to suffer. Suppose, for example, that while Curly is visiting a day-care center, he hears loud explosions that sound just like gunfire; sees a number of children screaming and crying, some of whom are lying on the ground and appear to be injured; and sees a man pointing a gun at them. When he yells at the man to stop, the man points the gun at one of the children, and Curly shoots him. If it turns out that the man's gun was a toy, the sounds were firecrackers next door, and the children were lying down and crying simply because they were tired, then Curly's act was clearly objectively wrong. He had no right to shoot the man who was

[13] A defender of desert-based retributivism might object to this assumption by maintaining that just and reasonable laws must take into account the moral beliefs of the people whose behavior they regulate. Just as the military draft, for example, makes an exception for conscientious objectors, it might be suggested, all just and reasonable laws would be written so that they do not apply to cases in which a person does something because she believes it to be morally right. If this suggestion were accepted, it would indeed rescue the desert-based retributivist position from this version of the undeserving offender objection. But the cost of doing so would clearly be too high: we would have to agree that just and reasonable laws would allow spousal rape and human sacrifice, for example, as long as the people committing these acts truly believed that they were morally permissible.

holding the toy gun. But since it seemed reasonable for Curly to believe that shooting the man was necessary to prevent him from killing innocent children, I doubt that anyone would think that Curly deserves to suffer for what he did. Yet, suppose now that Shemp, after thinking long and hard about the issue, believes that killing a doctor who performs abortions is necessary to prevent the doctor from killing many innocent children. Since it seems clear that Curly does not deserve to suffer for shooting the person he shot, and since it seems clear that the reason for this applies equally to the case of Shemp, I am not inclined to think that Shemp deserves to suffer even if we assume that his belief that abortion is murder is false. It is true, of course, that Larry and Curly act from a mistaken factual belief while Moe and Shemp act from a mistaken moral belief, but this difference in itself seems completely irrelevant to the issue at hand. As long as all four have taken equal care in forming their beliefs, and so are not blameworthy for their ignorance, it seems clear that they must be considered to be morally on a par. The lesson of such cases,[14] then, is that even when an illegal act is also immoral, it does not follow that the person who does the act deserves to suffer. If the person does the act from an unobjectionable or even admirable motive, then I, at least, see no force at all to the claim that the world would be a better place in a morally relevant sense if he suffers for what he did. And so, once again, the move from breaking a just and reasonable law to deserving to suffer proves unwarranted.

Defenders of the desert-based retributivist solution appeal to the intuitive reactions we are likely to have to the most extreme cases: cases in which the act is a very severe violation of the law, is a very immoral act, and the person who does the act is vile and unrepentant.[15] In such extreme cases, it is generally true that we will believe intuitively that the offender deserves to suffer. But this intuition will help to ground a solution to the problem of punishment only if it can support the more general claim that all persons who break just and reasonable laws deserve to suffer, even if their acts or motives are moral. Our intuition in extreme cases fails to support this more general claim, and our intuitions in other cases undermine it; so, in the end, the argument from cases fails to justify the desert-based retributivist position. The appeal to moral desert cannot account for all cases in which a defender of punishment believes that the state has the right to punish, and so the desert-based retributivist solution cannot pass the entailment test required for a satisfactory solution to the problem of punishment.

[14] Blanshard (1968) provides another useful example: that of the sincerely motivated traitor.

[15] This suggests a third problem case for the desert-based retributivist: one in which a person did an objectively immoral act for an objectively bad reason but is now genuinely repentant. I, at least, have difficulty feeling that this person now deserves to suffer, even if he might have deserved to suffer at the time of his offense.

3.1.3 The Punishing the Innocent Objection

Punishment involves punishing people for breaking the law. The person who breaks the law but does not deserve to be punished, therefore, presents a crucial problem for the desert-based retributivist defense of punishment. But punishment also involves not punishing people who do not break the law. And this points to a second problem for the desert-based retributivist. For just as there are people who break the law and do not deserve to suffer, there are surely people who do not break the law but do deserve to suffer. It is clear that on the desert-based retributivist account, the people we may permissibly punish are those who deserve punishment. As Moore puts it, "Moral culpability ("desert") is [according to desert-based retributivism] both a sufficient as well as a necessary condition of liability to punitive sanctions" (1987: 96). And as Kershnar puts it, "A person deserves punishment because, and only because, she has performed a culpable wrongdoing" [(2001: 41); see also (1999: 47)]. Since there are morally culpable wrongdoers who do not break the law, it follows that the desert-based retributivist solution is subject to the punishing the innocent objection. This means that the solution will again fail the entailment test and, since this result will also strike most people as seriously counterintuitive, it means that the solution will fail a reasonable application of the foundational test as well.

Imagine, for example, that the laws regarding spousal abuse are drawn up according to your specifications. They draw the line between behavior that is legally forbidden and behavior that is legally permitted precisely where you think it should justly and reasonably be drawn. Now consider a man who familiarizes himself with the law and does everything he can to make his wife miserable without crossing that line. If he is legally allowed to scream at her, he screams at her. If he is allowed to cheat on her, he cheats on her. In any way that he is allowed to embarrass, belittle, degrade, and insult her, he does, and with relish. He refrains from beating or raping her, but only because he is afraid of the legal consequences. Or consider the racist who does everything she is legally permitted to do to insult black people. If she is legally allowed to play racist songs, she does. If she is allowed to throw a party celebrating the fact that hundreds of black people died in an earthquake, she does. She refrains from burning crosses on people's lawns if she must, and she avoids lynching black people only because she is afraid of being caught. These people are legally innocent. They are immune to legal punishment. But if you have the intuition that desert-based retributivists want you to have in the kinds of cases that Moore and others typically cite, you will probably have the same sort of intuition here: these people deserve to suffer. Since they deserve to suffer, and since such desert is the basis for the right to punish on the desert-based retributivist account, it follows

that desert-based retributivism would render it morally permissible for the state to punish them even though they have violated no just and reasonable laws. After all, if the negative moral desert of the offender is sufficient to render it morally permissible for the state to harm him, then the equally negative moral desert of the nonoffender will be equally sufficient to render it morally permissible for the state to inflict an equal amount of harm on him. The offender who does not deserve to suffer and the nonoffender who does deserve to suffer are thus two sides of the same basic problem with the desert-based retributivist's appeal to our intuitive responses to particular cases. The basic problem is that the argument overgeneralizes from a relatively small and extreme set of cases.[16] And in doing so, it forces the desert-based retributivist to stray even further from the practice of punishment that he seeks to justify.

There are two ways in which a defender of the desert-based retributivist solution might respond to the punishing the innocent objection. One would be to deny that anyone who refrains from breaking the law is morally bad. This response is plainly implausible. Even if the laws are just and reasonable, there will be immoral behaviors that are not illegal. The other response would be to appeal to something other than the commonsense notion of moral badness when constructing the desert-based position in the first place. This, in the end, seems to be Moore's response to the problem. In a surprising and somewhat puzzling footnote, Moore explains that "'[m]oral culpability' as I am here using the phrase does not presuppose that the act done is morally bad, only that it is legally prohibited. An actor is culpable in this conception when, in doing an action violating some criminal prohibition, he or she satisfies those conditions of fair fault ascription" (1987: 96fn).

In one respect, of course, this stipulation would enable the desert-based retributivist to avoid the punishing the innocent objection (as well as the not punishing the guilty objection). Since by definition the nonoffender has not done a legally prohibited act, it follows that, on this conception of moral culpability, everyone who is legally innocent is morally innocent as well, and so does not deserve to suffer. But this

[16] This point has been noted by a number of writers, including, Bean (1981: 15), Lacey (1988: 19), and Dolinko (1991: 542). Stern, who discusses the issue in terms of "grossly immoral betrayals of friendship" of the sort that "no one wants the law to be concerned with," suggests that the problem can be overcome by appealing to the claim that since the betrayer deserves to be punished for his betrayal, such punishment would, in fact, "be justifiable if the impractical were practical" (1970: 323). But even if it were cost-effective for the state to punish people for every act of private immorality, it seems implausible to suppose that it would be permissible to do so, and agreeing that it would be permissible to do so would still be inconsistent with the theory's passing the entailment test since it would amount to admitting that the line between offender and nonoffender should not be treated as relevant in the way that legal punishment presupposes.

response to the punishing the innocent objection would save the desert-based retributivist from the objection only by depriving the position itself of the only support that initially grounded it. This is what makes Moore's stipulation here so difficult to understand. When we meditate on the kinds of cases that retributivists like Moore and Kershnar appeal to – those of vicious murderers, unrepentant rapists and torturers, and so on – we are not led to the thought that people who do *illegal* acts deserve to suffer. The Nazis so often cited to in this context, for example, did not break the laws of their country, and we would feel no less strongly about the just deserts of the rapist or murderer even if it turned out that their acts were legal. We are led, rather, to the thought that people who do *immoral* acts deserve to suffer. Indeed, Kershnar, at least, is explicit in recognizing this. He gives an example in which a woman points out to the Nazi authorities that they had overlooked a Jew who was hiding from them and says that his "intuition is that she deserves a rather severe punishment" even though, in this case, she acts in accordance with the law rather than against it (2001: 33).[17] And in discussing the ways in which our intuitions respond to the seriousness of an offense, he notes that "our intuitions tend to track proportionality, not in terms of the seriousness of the act as viewed by the law but rather the moral seriousness of the act" (2001: 83). This seems exactly right. But it is exactly what leads to the problem of the immoral nonoffender, who deserves to suffer as well, a problem that Moore seems to try to evade in a way that would undermine the only support he has provided for his position and that Kershnar seems to overlook entirely. And so, in the end, the desert-based retributivist is impaled on the horns of a dilemma: either the claim is that immoral acts merit suffering, which is well supported by the argument from cases but fails to justify legal punishment, or the claim is that illegal acts merit suffering, which helps to justify legal punishment but is not supported by the argument from cases. Either way, the desert-based retributivist solution will prove unable to pass the entailment test and either way, at least for most people, it will fail the foundational test as well.

3.1.4 The Act versus Outcome Objection

According to the desert-based retributivist position, an act of punishment is justified because, and only because, the person being punished deserves it. This position can justify punishment in general, therefore,

[17] See also Manser (1962: 301, 302). Relatedly, Kershnar gives an example of positive desert in which a good Samaritan saves a swimmer in shark-infested waters (2001: 10). Our intuition that this person deserves something good has nothing to do with the legal status of his act, only the moral status, and this provides still further support for the claim that the intuitions that desert-based retributivists appeal to track moral rather than legal wrongdoing.

only by showing punishment to be a kind of suffering that all offenders, and only offenders, deserve to undergo. This, I have argued, the desert-based retributivist cannot do. But let us now suppose that I have been mistaken, and that it is true that every offender and no nonoffender deserves to suffer. Even if this is true, the desert-based retributivist position must still be rejected for a further reason. There is an important gap between the claim that a person deserves something, on the one hand, and the claim that it is morally permissible to impose that something on the person, on the other. The first, as we have seen, amounts to the claim that the world is intrinsically a better place when she gets what she deserves than when she doesn't, while the second amounts to the claim that we have the right to force her to accept what she deserves whether she wants to or not. The desert-based retributivist position requires us to accept the inference from the former claim to the latter. But the inference itself is objectionable.[18]

That this inference is unacceptable can be seen in a number of ways. Certainly, at a general level, it does not immediately follow from the claim that one state of affairs is intrinsically better than another that we have the right to bring about the better state of affairs. The state of affairs in which only one innocent person dies is much better than the state of affairs in which five innocent people die, for example, but this does not in itself establish that it is morally permissible to kill one innocent person to prevent five others from dying.[19] Similarly, in the case of positive desert, the inference is plainly invalid. Suppose, for example, that five people will soon die if they do not get a new kidney, and only one new kidney is available. The five are physically equal, so there is no relevant difference in their compatibility with the new kidney. But they are morally unequal in the following sense: four of the five are mean, nasty, selfish people and the fifth is nice, friendly, and altruistic. In this case, it seems fair to conclude that the nice person is more deserving of the kidney – that the world in which he gets the kidney is a better world from the standpoint of desert than is the world in which one of the others gets it. But suppose that this person does not want to be given the kidney. Perhaps he thinks it should be used for research because in the long run that will do more good, or perhaps he opposes transplants on religious grounds. Clearly, in

[18] This problem has been noted, in one form or another, by a number of writers, including Barnett (1980: 142), Wolgast (1985: 167–70), Satre (1987–8: 432), Lacey (1988: 21–2), Ellis (1995: 227), Dimock (1997: 40–1), Hill (1999: 426), McDermott (2001: 405n2), and Golash (2005: 80).

[19] This is not to insist that it is impossible to defend the conclusion. One could defend it by defending a consequentialist moral theory, and so a consequentialist would be willing to accept the move from the goodness of a state of affairs to the rightness of bringing it about. But the fact that a consequentialist principle could be employed to support the inference is of no use to one seeking to defend punishment on retributivist grounds.

this case, even though the state of affairs in which he gets the kidney that he deserves is a better one, this does not justify the claim that we may coercively bring it about against his wishes.[20]

And now consider the offender who deserves to suffer. The state of affairs in which he suffers, that is, is better than the state of affairs in which he doesn't, not because of any further beneficial consequences that his suffering will bring about, but simply because deserved suffering in itself is better than the undeserved flourishing he will enjoy if he is not punished. And assume, furthermore, that this particular offender does not wish to be punished. In that case, to say that he deserves to be punished, and that this suffices to justify our forcing punishment on him, is on a par with saying that the nice patient deserves the kidney and that this suffices to justify our forcing the kidney on him. We surely will not say the latter, and so we cannot say the former. Moore, it should be noted, attempts to rebut this charge by appealing to the claim that those who deserve to be punished have forfeited their right not to be punished (e.g., 1993: 34, 36). If an offender deserves to be put in jail for five years, for example, and if she has forfeited the right not to be put in jail for five years by virtue of committing the offense, then the gap between deserving the punishment and the permissibility of inflicting it may be safely crossed. But if this is so, it will not be because desert-based retributivism has successfully overcome the act versus outcome objection. Rather, it will be because the retributivist will have abandoned desert as the justification for the permissibility of punishment in favor of the forfeiture of rights. And so, it is to the forfeiture-based version of retributivism that we turn next.

3.2 FORFEITURE-BASED RETRIBUTIVISM

The attempt to defend retributivism on desert-based grounds moves from our particular judgments about particular cases to a more general principle about desert and from there to a retributivist principle about punishment. I have argued that this strategy is unsuccessful. But the retributivist can also argue in the opposite direction: beginning with a

[20] Similarly, as Dolinko has pointed out (1991: 544), the fact that it would be better, from a desert standpoint, if a father's estate was passed on to his good son rather than to his wicked son does not make it permissible for the state to bring that about if the wicked son inherits the estate via his father's will [see also Dolinko (1997: 522–7)]. And more generally, as Moriarty (2003) has argued, it is difficult to avoid treating desert claims symmetrically in the context of distributive and retributive justice. But this means that, unless one believes that it is permissible for the state to coercively ensure that good people get the happiness they deserve, one cannot think it permissible for the state to coercively ensure that bad people get the unhappiness they deserve, even if one agrees that they deserve it and that things would be better if they got it.

broader theory about morality in general and deriving from it a retributivist principle about punishment in particular. Retributivists have primarily attempted to develop this approach in two ways: appealing on one version to a general theory about rights and on the other to a general theory about fairness.[21] I will focus in this section on the rights-based approach and in the following section on the fairness-based version. In the final section, it will address some less prominent further versions of the retributivist solution.

3.2.1 Punishment, Rights, and Duties

The attempt to derive a retributivist theory of punishment from a more general theory of rights begins by framing the problem of punishment in terms of rights. Punishment poses a problem, on this construal, not simply because it involves treating people in ways that would, under typical circumstances, be wrong, but because it involves treating people in ways that would, under typical circumstances, violate their rights. You have a right to life, for example, a right to control your property, and a right to freedom of movement. And so, at least in typical cases, it would be wrong for the state to take away your money, incarcerate you, or execute you because this would violate your rights. Yet, these are just the sorts of things that the state does when it punishes people for breaking the law. So, the problem of punishment becomes the problem of understanding how it can be morally permissible for the state to treat offenders in ways that typically involve violating people's rights.

[21] Corlett's book *Responsibility and Punishment* (2001) might at first seem to support retributivism by appealing to a third general theory, one that focuses on the notion of responsibility, but in the end, this notion seems only to express Corlett's belief in punishment rather than to support it. Corlett defines the liability sense of responsibility, so to say that a person is responsible for a certain action in this sense is to say that it would be appropriate to punish the person for the act (10), and given this fact, the claim that an offender is responsible in this sense for his offense cannot provide independent support for the claim that he should be punished. It can only be another way to say that he should be punished. Corlett could, of course, provide an independent argument to show that if an offender is responsible for breaking the law in the sense that he is at fault for breaking it, then he is also responsible in the sense that he may be punished for it. However, in the absence of such an argument, Corlett provides no reason to believe that one can move from a general theory of responsibility to a retributivist theory of punishment without simply assuming at the outset that punishment is morally permissible [which, at least at one point, he seems to do: "there is the *blame use of 'responsibility*,' that use of the expression which attributes accountability to those who are blameworthy for what they do. . . . It is assumed that the person who is responsible in the blame use of the term is one who, if certain other conditions are satisfied, is a candidate for moral censure and/or punishment and that they are at fault in what they did" (10, emphasis in the original)].

Once the question is framed in this way, the requirements for a successful solution become fairly clear. For punishment to be morally permissible, offenders must no longer have some of the rights that they had as nonoffenders. This, in turn, prompts a further question: why don't they have these rights? One possible answer appeals to the notion of consent. On this account, voluntarily committing a legal offense involves voluntarily consenting to give up some of your rights. This position is most closely identified with C. S. Nino (1983, 1991). Although it might seem reasonable to construe the consent-based defense of punishment as a version of retributivism, Nino explicitly declines to do so. For that reason, I will treat the consent-based position separately in Chapter 4. The other possible answer turns on the notion of forfeiture. Here the idea is that, by virtue of some voluntary act that you have done, you have lost or forfeited some of your rights even if there is no real sense in which you have agreed to give them up. The forfeiture-based retributivist position holds that the state has the right to punish offenders because, by having committed legal offenses, they no longer have the right not to be treated in the ways that punishment treats them. This view is often identified historically with Hobbes, Locke, and Hume[22] and has been defended more recently by a number of writers including Pilon (1978), Goldman (1979), Rothbard (1982), Haksar (1986), and Kershnar (2002).

The argument for the forfeiture-based retributivist solution begins with what I will refer to as the "forfeiture claim," the claim that if *P* violates *Q*'s right to *X*, then *P* forfeits *P*'s own right to *X* (or perhaps instead forfeits some equivalent right or set of rights).[23] To this foundational claim, the argument adds the plausible assumption that offenders violate the rights of others. If both of these claims are true, it follows that offenders forfeit some of their rights in a way that nonoffenders do not. As Goldman puts it, "by violating the rights of others in their criminal activities, [offenders] have lost or forfeited their legitimate demands that others honor all their formerly held rights" (1979: 31). The fact that a particular offender has forfeited a particular right, of course, does not in itself give the state a good reason to harm him in ways that would formerly have violated that right. On Goldman's account, at least, punishment will still be a good idea only if it serves some useful purpose.[24] But the central point of the forfeiture-based retributivist position

[22] Locke's defense of forfeiture-based retributivism is scattered throughout the First and Second Treatises. For a useful discussion of this, see Simmons (1991: 238ff.). For the view of Hobbes and Hume as proponents of the forfeiture-based position, see Morris (1991).

[23] As Goldman, e.g., puts it "if we ask which rights are forfeited in violating rights of others, it is plausible to answer just those rights that one violates (or an equivalent set)" (1979: 33).

[24] "When a person violates rights of others, he involuntarily loses certain of his own rights, and the community acquires the right to impose a punishment, *if there is a social benefit to be derived from doing so*" [Goldman 1995: 32, emphasis added].

is that serving a useful purpose is not enough to make it permissible to punish an offender in the first place were it not for the fact that he had forfeited some of his previous rights by committing the offense for which he is to be punished. We have the right to punish such offenders because, and only because, they have lost some of their previous rights. Or, to quote Goldman once more: "One continues to enjoy rights only as long as one respects those rights in others: violation constitutes forfeiture" (1979: 33).

The argument for the forfeiture-based retributivist solution begins with the forfeiture claim: if *P* violates *Q*'s right to *X*, then *P* forfeits *P*'s own right to *X* (or perhaps some equivalent right or set off rights). The assessment of the forfeiture-based retributivist solution should therefore begin by asking why we should take this claim to be true. In response to this question, the forfeiture-based retributivist has two options: she can simply assert that the claim is true, or she can give an argument for it. Some defenders of the forfeiture-based retributivist position seem to be content with the first option. And since the forfeiture claim is likely to strike many people as plausible, this may initially seem to be a reasonable response. But it is not. The forfeiture claim is hardly self-evident. Indeed, on many of the most common views of rights, a person has the rights that he has in virtue of some essential property of his (sentience, rationality, autonomy, humanity, etc.); as a result, it is difficult, if not impossible, to see how he could lose them at all, let alone lose them simply by violating someone else's rights. If there is no strong, independent reason to believe that offenders forfeit their rights, then the only reason for saying that they do is that saying so will lead the retributivist to the desired conclusion that punishment is morally permissible. So, it seems clear that a defense of retributivism on forfeiture-based lines must provide some independent argument for the claim that if *P* violates *Q*'s right to *X*, then *P* forfeits *P*'s own right to *X* (or some equivalent right or set of rights).

A number of retributivists have attempted to provide just such an argument. The most prominent argument by far is based on the claim that there is an important relation between moral rights and moral duties.[25] If, in particular, affirming moral rights involves affirming moral

[25] The argument based on this claim is defended by Pilon (1978: 355–6), Goldman (1979) and Haskar (1986: 321ff.) [For a useful critique of Haskar, see Brady (1987)]. An argument for the claim that offenders forfeit rights might also be developed by appealing to the claim that it would be unfair to allow rights violators to enjoy the same rights enjoyed by nonviolators [see, e.g., Simmons (1991: 243–4)]. This position will be treated as a distinct fairness-based version of retributivism in Section 3.3, but all of the objections raised against what I refer to there as fairness-based retributivism would apply equally to it if it were reformulated as a version of forfeiture-based retributivism. McDermott has offered a further argument for what he sometimes characterizes as a forfeiture version of retributivism, an argument based on the notion of debt payment (e.g., 2001: 424, 426).

duties, then it would seem to follow that negating one's moral duties would involve negating one's moral rights. As Goldman puts it, "Since having rights generally entails having duties to honor the same rights of others, it is plausible that when these duties are not fulfilled, the rights cease to exist" (1979: 31). I will refer to this as the "rights-duties argument." On the face of it, at least, the rights-duties argument helps to make forfeiture-based retributivism an attractive position. It grounds the forfeiture claim in a more general view about rights and duties, does so in a manner that does not seem to be objectionably ad hoc, and provides the forfeiture-based retributivist solution with a seemingly airtight defense against the problem of punishing innocent people: since only those who commit offenses forfeit their rights, and since only those who forfeit their rights may permissibly be punished, only those who commit offenses may be punished. While forfeiture-based retributivism, on this account, may therefore represent an improvement over the desert-based alternative, however, it is in the end undermined by a variety of problems.

3.2.2 The Rights Without Duties Objection

The first problem lies with the forfeiture-based retributivist's attempt to defend the forfeiture claim by appealing to the further claim that rights entail duties. And the first problem with this attempt is that the further claim that it appeals to seems to be false. Very young children and infants, fetuses, nonhuman animals, and people with severe mental disorders, for example, are not moral agents. They are not the kinds of beings that can be held morally responsible for their behavior. They therefore have no moral duties. If the claim about the relation between rights and duties appealed to in defending the forfeiture claim were correct, it would follow that these individuals also have no rights. But while people may agree that at least some of these individuals have no rights, it is extremely implausible to suppose that none of them do. And even in the case of those who, we might agree, lack rights, it is difficult to accept that their mere lack of moral agency would suffice to establish this. Many people agree that a fetus does not have a right to life, for example, but it is implausible to suppose that the entire debate about the rights of the fetus could be resolved simply by pointing out that fetuses do not have duties.

Since his argument differs in important ways from the forfeiture-based position considered here, it, too, is treated separately in Section 3.4.2. Finally, Morris (1991) has offered a defense of the forfeiture-based position that is closely related, but not identical, to the rights-duties argument outlined in this section. While Morris's version of the forfeiture-based retributivist position may prove immune to some of the objections that I raise against the rights-duties argument in particular (in Sections 3.2.2 and 3.2.3), it remains susceptible to those objections that I raise directly against the forfeiture-based solution itself (in Sections 3.2.4–3.2.8).

The claim that rights entail duties serves as the foundation of the argument for the forfeiture claim, which in turn serves as the foundation for the forfeiture-based retributivist solution. So, if that claim is defeated, the entire forfeiture-based retributivist position collapses.

A defender of the rights-duties argument for the forfeiture claim can, of course, respond by restricting the scope of the claim that rights entail duties. Goldman, for example, at one point puts the argument's foundational claim as follows: "a condition of having specific rights is that one honors those rights of others (*when one is able to do so*)" (1979: 32, emphasis added). On this account, the duty not to violate a particular right of a particular individual is simply a duty to refrain from violating that right when one is able to do so. Since an infant, for example, lacks the abilities that would enable it to refrain from violating someone else's rights, it would therefore follow from this qualification that an infant can, in fact, have a duty not to violate the rights of others. And since it can have this duty, it can have the corresponding rights.

But this stipulation is objectionable for two reasons. First, it is ad hoc. Our natural inclination is to say that children acquire moral obligations only when they reach a certain level of maturity. The forfeiture-based retributivist must provide a reason for making our view more complicated than this, and the reason must be other than the fact that doing so will enable her to avoid a problematic implication of her position. Second, and more importantly, this stipulation would render vacuous the claim that everyone who has rights also has duties. After all, if having a duty includes the stipulation that one must be able to exercise the duty, then everything that doesn't have rights will have duties, too. We would have to say, for example, that a grain of sand has a duty not to break its promises, since if it had the ability to make and break promises, it would be wrong to break them. I therefore conclude that the assumption needed to ground the forfeiture claim in the relationship between rights and duties is unwarranted and that it gives us no reason to accept the forfeiture claim. And if this is so, then the forfeiture-based retributivist solution will clearly fail the foundational test that any solution to the problem of punishment must pass.

3.2.3 The Rights Without Fulfilling Duties Objection

I have argued that the claim about rights and duties used to underwrite the forfeiture claim is false. But let us now suppose that it is true. Having a particular right really does entail a duty to respect that right in others. What follows from this? According to the rights-duties argument, what follows is that a person who violates the duty to respect a particular right in others does not have that particular right himself. But this does not follow at all. What follows is simply that a person who does not *have* a duty

to respect a particular right in others does not have that particular right himself. The claim that having *A* entails having *B*, after all, merely justifies the claim that not having *B* entails not having *A*. And this conclusion is not sufficient for the forfeiture-based retributivist's purposes. A murderer, for example, certainly violates her duty not to kill others, but this does not mean that she does not *have* such a duty. And as long as she still has such a duty, the fact that having rights entails having duties cannot be used to show that by violating her duty, she has lost her right to life.

Now in fairness to Goldman, it is important to emphasize that he does not insist that the claim that rights entail duties logically entails the forfeiture claim. He says only that acceptance of the first claim makes the second one "plausible." But the very reason for concluding that the move from the former to the latter is logically invalid is at the same time a reason for concluding that the move is implausible as well. What makes the inference invalid, after all, is the fact that violating a duty is consistent with having that duty. Indeed, violating a duty positively *requires* that one have that duty, since if one does not have the duty in the first place, one cannot do anything to violate it. But if violating a duty to respect a particular right means that one has a duty to respect a particular right, and if rights really do entail duties, then this would, if anything, seem to suggest that since a person who has violated his duty still has his duty, he still has his rights. The fact that he still has his duty does not prove that he still has his rights, of course, since the initial claim was only that rights entail duties, not that duties entail rights. But if the argument is to be assessed at the level of surface plausibility rather than deductive validity, it seems at least as fair to say that the supposed connection between rights and duties makes plausible the view that an offender who still has a duty to respect the rights of others still has that right against others. In any event, the move from the claim that rights entail duties to the forfeiture claim looks plausible only because it looks valid. Once we come to see that it is not valid, it does not seem plausible either. Even if the claim that rights entail duties is accepted, therefore, the argument for the forfeiture claim must still be rejected and, once again, the forfeiture-based retributivist solution must be judged to have failed the foundational test. Whether or not it would entail the permissibility of legal punishment, the principle of forfeiture it appeals to is insufficiently warranted on its own terms.

3.2.4 The Unforfeited Rights Objection

I have maintained that the primary argument thus far provided for the forfeiture claim should be rejected. The rights-duty argument's premise is false, and even if it were true, it would fail to support the conclusion that the forfeiture claim is true. It must be conceded, however, that the claim that those who violate a particular right forfeit that right has a certain

intuitive plausibility for many people. It is therefore important to see that there are several additional reasons to reject the forfeiture-based retributivist solution, even if the basic idea behind the forfeiture claim at first seems to be attractive.

The first reason arises from the fact that even though the forfeiture claim seems plausible to many people in the case of some particular rights, it seems to be extremely implausible in the case of many others. These other cases provide powerful counterexamples to the forfeiture claim and thus constitute further reasons to conclude that the forfeiture-based position ultimately fails a reasonable application of the foundational test. One case concerns rights against certain violations of bodily autonomy. The forfeiture claim, for example, entails that a rapist forfeits the right not to be raped and a torturer forfeits the right not to be tortured. A second case involves what might be called "intellectual rights," the rights that a government official would violate, for example, if he prevented members of a particular religious denomination from gathering to worship or if he confiscated and destroyed a reporter's notes before she was able to produce a story based on them. The forfeiture claim would entail that in such cases the official in question has lost his own right to religious freedom or to freedom of expression. A third case turns on what might best be described as "procedural rights." A person who has been convicted of breaking the law, for example, presumably has a right to have his sentence determined in a fair and impartial manner. A judge who takes a bribe and as a result hands down an unfair sentence violates this right of the offender. The forfeiture claim would therefore entail that if the judge is convicted of this offence, she has forfeited her own right to receive a fair sentence.[26] It is difficult to believe that these people have lost these rights: that it would be morally permissible for the state to torture the torturer, censor the censoring official, unfairly sentence the unfair judge, and so on. If we are unwilling to believe these things, then we must be unwilling to accept the forfeiture claim and must again conclude that the forfeiture-based retributivist solution fails the foundational test.

Some retributivists, of course, may be willing to bite the bullet at this point and agree that all of these forms of punishment would be morally permissible. Most, I suspect, would not. These retributivists would be forced to concede that at least for some rights, perhaps because they are inalienable or for some other reason strongly connected to the nature of the person who bears them, the rights remain unforfeited even after the

[26] My discussion of the second and third counterexamples to the forfeiture claim borrows significantly from Burgh (1982: 198). Lippke also presses the first case against the forfeiture-based retributivist (1998a: 538; 1998b: 31), and Braithwaite and Pettit also appeal to the second case (1990: 169).

offender has violated them in someone else. But if this is so, then these particular offenders will have lost no rights as a result of their wrongdoing, so there will be no basis for maintaining that it is permissible for the state to punish them.[27] For these retributivists, then, the forfeiture claim must be modified in a way that allows it to avoid both the implication that it is permissible to torture the torturer and the implication that it is permissible to do nothing to him at all.

At first, it might seem that all of this could be accomplished simply by appealing to the notion of "equivalent" rights. Goldman, for example, suggests that "if we ask which rights are forfeited in violating rights of others, it is plausible to answer just those rights that one violates (*or an equivalent set*)" (1979: 33, emphasis added). On this account, the forfeiture-based retributivist might maintain that the torturer, for example, still retains his right not to be tortured, but has instead lost some other right or set of rights that are, in some sense, equivalent to that right. And assuming that the resulting loss of these other rights does not strike us as intuitively unacceptable, this modification will permit the forfeiture-based retributivist to overcome the problem posed by cases involving unforfeited rights.

The reason to reject the equivalent rights response to the unforfeited rights objection is simple: if a given right really is equivalent to another, then if it is unacceptable to deprive someone of one right, it must be equally unacceptable to deprive him of the other. If, on the other hand, we have no qualms about depriving someone of a certain right but would strongly resist depriving him of some other right, then this fact in itself demonstrates that the rights are not equivalent. Even if we feel that we can identify a right or set of rights that is equivalent to the rights that an offender has violated, therefore, this will do nothing to overcome the unforfeited rights objection.

This problem with the equivalent rights response is perhaps most clearly seen by applying it to a particular set of cases. Suppose, to take perhaps the most widely accepted punishment, we focus on incarceration. Imprisoning an innocent person presumably violates her right not to be imprisoned, and the longer the period of imprisonment, the greater the violation of this right. So, a proponent of the equivalent rights response might attempt to overcome the problem posed by such cases as those of the rapist and torturer, the censoring official and the corrupt judge by

[27] This problem will also arise when an offender violates a right that his victim has but that he does not have. Suppose, for example, that a woman has a right to an abortion and a man violates this right by preventing her from getting it. The forfeiture claim would seem to entail that the man must thereby forfeit his own right to an abortion, but since he has no such right to begin with, it is once again difficult to see how the forfeiture claim can produce the results desired by the retributivist.

finding a way to convert the rights violations that each of them inflicted on their victims into a right not to be imprisoned for a particular period of time. Lippke, for example, proposes a retributivist method for determining appropriate prison sentences. The sentence is appropriate if "the effects of punishment interfere commensurately with the capabilities of offenders to live decent lives of their own choosing" (2003: 34–5). So, if being raped interferes to some degree with a victim's ability to live a decent life of her choosing and if being censored interferes with a victim's ability to live a decent life of her choosing to some other degree, then the forfeiture-based retributivist could say that the rapist and the censor have each forfeited their right not to be imprisoned, and that the former has lost the right to one degree and the latter to another.

When placed in this context, the problem with the equivalent rights response can be put as follows: suppose we agree that putting a torturer in prison for a certain number of years will interfere with his ability to live a decent life of his own choosing just as much as he interfered with his victim's ability to do so by torturing him. If this is true, and if we are unwilling to say that he has lost his right not to be tortured, then why should we be willing to allow that he has lost his right not to be imprisoned? Proponents of the equivalent rights response have provided no answer to this question. And until it can be answered, the unforfeited rights objection should be accepted as a further reason to deny that the forfeiture-based retributivist solution can pass a reasonable application of the foundational test.

3.2.5 The Disproportionate Punishment Objection

A second problem with resting a defense of retributivism on the seeming plausibility of the forfeiture claim arises from an ambiguity in the forfeiture claim itself: when we say that an offender forfeits the right that she violates in her victim, should this be understood to mean that she forfeits the right permanently or only temporarily? In some cases, this question may prove relatively simple to answer. Since the murderer permanently violates the right to life of her victim, for example, it would seem relatively straightforward to say that she has permanently forfeited her own right to life. But most offenses seem to involve only a temporary violation of the victim's rights, and in many of these cases, the forfeiture claim cannot be so easily interpreted. Whether the offender is said to lose her right permanently or only temporarily, the result of using the forfeiture claim as a basis for the right to punish her will be forms of punishment that are objectionably disproportionate to the gravity of the offense. For anyone committed to at least a rough proportionality between punishment and offense, therefore, this problem with the forfeiture claim will provide yet another reason to conclude that the forfeiture-based retributivist solution fails the foundational test.

Consider, for example, the case of kidnapping. Suppose that Larry kidnaps Moe and holds him against his will for three days in a room of the same size, and with the same amenities, as a standard prison cell. Suppose, moreover, that in doing so he gets Moe to (correctly) believe that he is in no danger of being physically harmed and that he will be safely released at the end of the three days. At the end of the three days, Larry lets Moe go. What does the forfeiture claim entail about such a case? Larry has clearly violated an important right of Moe's: the right to freedom of movement. If we accept the claim that the violation of a particular right entails the forfeiture of that right, then we must conclude that Larry has forfeited his right to freedom of movement. But has he forfeited this right temporarily or permanently? There are two possible answers to this question. Neither can help the forfeiture-based retributivist avoid the disproportionate punishment objection.

Suppose first that the forfeiture-based retributivist answers that Larry has permanently lost his right to freedom of movement. In that case, it will be morally permissible to imprison Larry for the rest of his life as a result of his offense. But this seems clearly to be too drastic. Furthermore, it would seem to imply that every person who violates someone else's freedom of movement could be put in prison for life, regardless of how long he confined his victim. Even an offender who locked someone in a closet for only fifteen minutes or fifteen seconds could receive a life sentence. This is clearly too much punishment. But suppose instead that the forfeiture-based retributivist responds by saying that Larry has lost his right to freedom of movement only temporarily. The question that then arises is: for how long? There would seem to be only one nonarbitrary answer to this question. Since Larry violated Moe's right for three days, he should be understood as having lost this right for three days. But on this answer, the state is entitled to imprison Larry for only three days. And intuitively, this seems to be too little.[28]

The forfeiture-based retributivist, of course, can say that we can deprive Larry either of three days of freedom of movement or of an equivalent amount of some other freedom, but this response is clearly inadequate. If depriving Larry of only three days of freedom of movement is far too little punishment for his offense, then so too will be depriving him of something else that is equivalent to it. Yet, if the forfeiture-based retributivist wants to say that Larry has forfeited his right to freedom of movement for more than three days but less than the remainder of his life, there seems to be no nonarbitrary means of determining how long that would be without simply abandoning the foundations of the forfeiture-based position. We might believe, for example, that Larry deserves to spend at least six months in prison. And

[28] This problem is also noted by Lippke (2001: 82).

so, if the forfeiture-based retributivist maintained that by violating Moe's right to freedom of movement for three days he had forfeited his own right to freedom of movement for six months, then the result of applying the forfeiture claim might strike us as intuitively acceptable. But this would be so only because we had relied on considerations of desert rather than considerations of rights in determining how much punishment it would be permissible to inflict on Larry. It would do nothing to show how considerations of rights could be led to yield an intuitively acceptable level of punishment. And so, adhering to forfeiture-based retributivism in this case would mean accepting either too much punishment or too little. And this problem, it should be emphasized, will arise not just in the case of kidnapping, but in the case of many other important offenses as well. An offender who punches someone in the nose, steals a car for a brief joyride and then returns it, or illegally listens in on one telephone call violates his victim's rights for a relatively short period of time. So, in all of these cases, and many more as well, the forfeiture-based retributivist will have to say either that the offender has forfeited his corresponding rights permanently (resulting in too much punishment) or very briefly (resulting in too little punishment). Since most people who defend punishment do so only if they can defend proportionate punishment, this fact provides a further reason to conclude that the forfeiture-based retributivist solution fails a reasonable application of the foundational test.

3.2.6 The Private Retaliation Objection

A third worrisome implication of forfeiture-based retributivism concerns violence against offenders by private citizens. Suppose, to begin with an extreme case, that a convicted murderer has forfeited his right to life. The fact that he no longer has a right to life is meant to explain why it is morally permissible for the state to kill him. But if it is true that he no longer has a right to life, then this fact should make it equally morally permissible for anyone to kill him.[29] This would apply both to people who wish to kill him because he committed a murder and to people who wish to kill him because they don't like him or because they simply enjoy killing other human beings. And the same would be true of lesser offenses: whatever harmful treatment the state inflicted on an offender would be justified by the fact that the offender no longer had a moral right against being subject to such treatment. And if the offender no longer has such a right, then it would be morally permissible for anyone

[29] A defender of forfeiture-based retributivism could, of course, insist that the murderer only forfeits his right to life with respect to the state, but this response seems to be entirely ad hoc. When the murderer killed his victim, after all, he did not violate the murderer's right not to be killed by the state, but simply his right not to be killed.

to inflict comparable punishment on him.[30] Some defenders of punishment may not flinch at this implication. Many others, however, will. What they seek to defend is not merely the claim that it is permissible for the state to punish, but the claim that it is permissible for the state to maintain a monopoly on punishment. For them, at least, as well as for anyone else who is disturbed by the prospect of permissible private violence against offenders, this implication should provide another reason to conclude that the forfeiture-based solution fails the foundational test.

3.2.7 The Punishing the Innocent Objection

I have identified a number of reasons to deny that the forfeiture-based retributivist solution can pass the foundational test. The forfeiture claim that the solution depends on is implausible; the claim about rights and duties used to support the forfeiture claim is itself implausible and fails, in any event, to support it; and the forfeiture claim has a variety of implications that are themselves objectionable and that provide still further reason to reject the forfeiture approach. But let us now suppose that I have been mistaken about all of this and that the foundation of the forfeiture-based retributivist solution is sound and acceptable. Even if all of this is true, the solution must still be rejected for a further reason: it fails to justify the practice of punishment. There are two reasons to conclude that even if the forfeiture-based retributivist solution can pass the foundational test, it will fail the entailment test.[31]

The first reason arises from a further ambiguity in the forfeiture claim itself: should the notion of forfeiture that it appeals to be understood in terms of moral rights or legal rights? Suppose first that the claim is cashed out in terms of legal rights. When you violate a victim's legal rights, on this version of the claim, you forfeit your corresponding legal rights. Every citizen has a legally protected right to life, for example; so, when a murderer violates his victim's legal right to life, he forfeits his own legal

[30] Simmons suggests that the forfeiture-based retributivist can respond to the case of the person who kills the murderer for some other reason by changing the position to one in which the murderer has forfeited only his right not to be killed for certain reasons (1991: 248). This response, too, strikes me as unacceptably ad hoc. Although Simmons is correct to point out that we sometimes voluntarily transfer rights that are qualified in this way, it is difficult to see why this would be so in the case of forfeiture. After all, when a murderer kills his victim, he does not simply violate his victim's right not to be killed for certain reasons. He violates his right to life. Further difficulties with Simmons's response are identified by Lippke (2001: 81–2).

[31] There may be three such reasons. In addition to the final two objections I raise in this section, it is not clear that the forfeiture-based retributivist solution can escape the not punishing the guilty objection. This is because, in a number of cases, such as building an addition to one's home without obtaining the required permit or driving (safely) without a license, it is not clear whose rights, if any, an offender violates in the first place.

right to life. If this is the position of the forfeiture claim, then it cannot underwrite a solution to the problem of punishment. For the claim that a murderer has no legal right to life is perfectly consistent with the claim that he has a moral right to life. And if he has a moral right to life, then executing him, although legal, is immoral. But the problem of punishment is not the problem of figuring out how legal punishment can be legal. It is the problem of figuring out how legal punishment can be moral. And so, on this first construal, the forfeiture claim will prove irrelevant to solving the problem of punishment. It will entail nothing at all about the moral permissibility of punishing anyone and will thus render the forfeiture-based retributivist solution unable to pass the entailment test.

But suppose, on the other hand, that the forfeiture claim is understood in terms of moral rights. Every citizen has a moral right to life, for example; so, when a murderer violates his victim's moral right to life, he forfeits his own. If the forfeiture claim is understood in this way, then it can explain why it is morally permissible (and not simply legally permissible) to execute the murderer. But, unfortunately for the forfeiture-based retributivist, when the forfeiture claim is understood in this way, it will also prove much more than this. In particular, it will prove that it is also morally permissible for the state to punish people who are morally guilty but legally innocent. It will prove this because not every moral right is, or should be, protected by a corresponding legal right. A person who violates the rights of another when the violated right is a moral but not a legal right does something that is immoral but not illegal. And so, the person in such cases is, in the relevant sense of the word, innocent. But since such persons have violated someone else's moral rights, it follows from the forfeiture claim on this second interpretation that they have forfeited their own corresponding moral rights. And from this, it follows that it would be morally permissible for the state to punish them even though they have done nothing illegal. The second possible construal of the forfeiture claim, in short, will subject the forfeiture-based retributivist solution to the punishing the innocent objection, and since punishment treats the line between offender and nonoffender in a manner that is incompatible with this objection, the solution will again prove incapable of passing the entailment test. Even if the forfeiture claim so construed is accepted, then, what it will justify is not the practice of punishment.

The problem of punishing the legally innocent will arise for the forfeiture-based retributivist on this understanding in a variety of contexts. Consider, for example, the moral right not to be lied to or not to have a promise broken. In the context of such offenses as perjury and fraud, it seems just and reasonable to prohibit acts that violate these rights. But in other contexts, it seems equally clear that just and reasonable laws would permit such violations. Suppose, for example, that Larry is dating

Betty and promises her that he will not see anyone else. Larry then goes behind Betty's back, has an affair with Betty's best friend, and lies to Betty about it afterward. Morally speaking, Betty has a right to be told the truth and to have the promise that was made to her kept. Larry has violated these rights. And so, if the forfeiture claim is understood in terms of moral rights, it follows that Larry has forfeited these moral rights or some equivalent set of them. And from this, it follows that it is permissible for the state to punish Larry even though he has violated no just and reasonable law. Or consider the moral right not to be physically harmed or to be secure in one's property. In general, it again seems clear that just and reasonable laws would protect people from violation of these rights. But, again, there seem clearly to be exceptions: cases in which an act would violate the moral right without violating a just and reasonable law. Consider, for example, cases in which the harm done is relatively mild. During a quarrel, for example, Betty slaps Larry in the face, or while Moe and Curly are eating lunch at a restaurant, Moe takes a few of Curly's potato chips without first getting his permission. In these cases, an act is done that violates the moral rights of its victim. If the forfeiture claim is construed in terms of moral rights, as it must be to prove relevant to the problem of punishment, then it will again entail that it would be morally permissible for the state to punish people who are innocent of any violation of the law. And this will prevent the forfeiture-based solution from passing the entailment test (as well as preventing it, for many people, from passing the foundational test for yet another reason). Whether the rights in question are construed as moral rights or as legal rights, therefore, the forfeiture-based retributivist solution cannot pass the entailment test.

3.2.8 The Harm versus Punishment Objection

There is a second and very different reason to conclude that the forfeiture-based solution cannot pass the entailment test. This reason assumes, at least for the sake of the argument, that all of my previous objections have been overcome and that the forfeiture claim implies that all and only offenders would lose some of their moral rights and would do so to a proportionate degree. This further problem maintains that even if all of this is so, the practice of harming those who had violated the law that the forfeiture claim would justify is not the practice of punishment.

To see how this final problem for the forfeiture-based retributivist solution arises, it is helpful to begin with a right whose alienability is uncontroversial and a context in which the conditions that constitute forfeiting that right are also uncontroversial. So, consider the case in which you have leased a car from a dealership and have signed a contract agreeing to pay the dealership $300 a month for the next three years.

Assuming that you have kept up with your payments, you have a right to use this car. Further, this right is alienable. If you wish to return the car to the dealership, for example, you may do so; at that point, you will no longer have the right to use it. Furthermore, it is clear what would count as forfeiting your right to use the car: your failure to make your monthly payments.

Now, suppose that you do fail to make your monthly payments, and the dealership demands that you return the car. And suppose that, for whatever reason, you refuse to do so. The dealership goes to court and the judge orders you to return the car. In issuing this ruling, the judge is doing something to you that she could not do if you had not forfeited your right to use the car. If you still had that right, the judge could not now take the car from you. In addition, in forcing you to give up the car, the judge is doing something that harms you, and she recognizes that this harms you. But – and this is the crucial point – the judge is not taking the car away from you with the *intention* of causing you harm and making you suffer. She is merely taking the car away from you in order to make sure that you do not keep something you have no right to possess. She foresees that in enforcing the terms of the contract you will be harmed, but harming you is not her intention. And because of this important fact, what the judge does to you when she takes the car cannot be classified as punishment. As we saw in some detail in Section 1.1.4, after all, punishment involves not just harm but intentional harm. A fine is a form of punishment, for example, while a fee is not.

But now consider that, on the forfeiture-based retributivist account, what the judge does to you because you have lost your right to use the car is precisely what any judge would do to any offender who had lost any right. Suppose, for example, that it makes sense to say that an arsonist has forfeited his freedom of movement for five years. And suppose, in addition, that because he has lost five years' worth of freedom of movement, a judge is now entitled to put him in prison for five years. In that case, the judge is entitled to take away the arsonist's freedom to deprive the arsonist of something to which he is no longer entitled, and the judge is entitled to do so with the understanding that this will harm the arsonist. But since, on the forfeiture-based retributivist account, the case of the arsonist is parallel to the case of the car lease, and since the judge does not have the right to deprive you of the car as a means of harming you, it follows that the judge in the arson case is not entitled to imprison the arsonist with the intention of harming him either. And so, in this case, just as in the case of the leased car, the arsonist is not punished. He is harmed, certainly, but he is not punished. To have something taken from you because you are not entitled to have it in the first place is not to punish you.

From a practical point of view, of course, this final objection is essentially irrelevant. If the forfeiture-based retributivist can justify imprisoning

someone for committing arson, after all, what does it matter if we call imprisoning the arsonist punishment or something else? But from a theoretical point of view, this final objection is decisive. The problem of punishment, after all, arose in the first place because punishment involves intentionally harming people and because intentionally harming people is, in general, morally impermissible. Most people believe that the intentional infliction of harm, though wrong in general, is justified in the case of people who have broken the law. The problem of punishment is the problem of finding a principled justification for this belief. The harm versus punishment objection demonstrates that the forfeiture-based retributivist solution cannot justify this belief because it cannot justify the intentional infliction of harm. And if it cannot justify the intentional infliction of harm, then even if it can overcome all of the other objections raised against it here, it still cannot solve the problem of punishment. There are a number of good reasons to deny that the forfeiture-based retributivist solution passes a reasonable application of the foundational test, therefore, but even if it does, it cannot pass the entailment test. Either way, it does not solve the problem of punishment.

3.3 FAIRNESS-BASED RETRIBUTIVISM

Forfeiture-based retributivism attempts to justify punishment by grounding retributivism in a general theory of rights. I have argued that this attempt is unsuccessful. But the retributivist can attempt to justify punishment by grounding retributivism in a general theory of some other sort. In particular, a number of retributivists have attempted to ground retributivism in a general principle of fairness. This fairness-based version of retributivism is perhaps most widely associated with Herbert Morris's influential paper "Persons and Punishment" (1968) and George Sher's widely discussed book *Desert* (1987), but it has also been defended in one form or another by a number of recent writers including Finnis (1972; 1980: 261–6; 1999: 98–103), Dagger (1993), and Murphy (1973), and, at least on some interpretations, was defended much earlier by such historical figures as Kant and Hegel.[32] There is little question, then, that

[32] An earlier version of Morris's position is also briefly suggested in (1965: 373–5). Murphy defends a version of fairness-based retributivism in (1973) but later expresses doubts about it (e.g., in 1990c). Sher responds to some of the criticisms of his 1987 book in (1997). In addition to the writers cited previously, the position has been defended by Gerstein (1974: 76–7), Gewirth (1978: 294–8), Sadurski (1985: chap. 8; 1989), von Hirsch (1976: chap. 6), Bradley (1999: 106–9ff.), and (with an emphasis on the duty and not merely the right to punish) McDermott (1999: 151–63). In addition, Adler has defended a version of retributivism that is cashed out in terms of equality, but the core of the argument seems to be essentially the same as the argument considered here [see, e.g., (1988: 255–61; 1991: 121–5)]. Doyle (1967) can also be seen as a precursor to this view,

the appeal to fairness is a popular alternative to the forms of retributivism already considered in this chapter. Indeed, the fairness-based approach is arguably the preeminent form of retributivism in the current literature. The question, then, is whether it is more satisfactory than its rivals. I will argue in this section that it is not.

3.3.1 Punishment and Fair Play

While the forfeiture-based retributivist solution begins by framing the problem of punishment in terms of rights, the fairness-based retributivist solution begins by framing it in terms of distributive justice. Punishment, that is, involves imposing certain burdens on some people that are not imposed on others. The problem of punishment, on this account, involves explaining how it can be fair to distribute these burdens so inequitably. Once the problem of punishment is framed in this way, the general shape of the answer that the fairness-based retributivist must provide becomes clear: offenders must be viewed as enjoying an unfair distribution of benefits as a result of having committed an offense, so that imposing on them a punitive harm will restore the overall distribution of benefits and burdens to its previous and presumptively fair level. The problem for the fairness-based retributivist then involves filling in the details that would make this analysis true.

The filling in begins with the endorsement of a general principle of fairness that has come to be known as the "principle of fair play." The principle, in Dagger's formulation, maintains that "anyone who takes part in a cooperative practice and accepts the benefits it provides is obligated to bear his or her share of the burdens of the practice" (1993: 475). To accept collectively produced benefits without incurring the costs that others incur in producing them is to be a free rider, and the principle of fair play can therefore be represented as the claim that it is unfair to be a free rider. To this general and quite plausible moral view, the fairness-based retributivist adds a basic descriptive claim about the nature of society in general. The claim is that society is best understood as a mutually beneficial venture made possible by mutual cooperation. Each person benefits from living in a world of order rather than disorder, that is, but such order exists only because people generally abide by the rules

and Sterba (1977, 1990) defends the view, though from a rational choice rather than a straightforwardly moral vantage point (see esp. 1990: 176–7). Day (1978: 505–6) offers an argument from fairness to what he calls "retributive punishment," though his conclusion seems instead to involve compulsory victim restitution. For the claim that Kant was a proponent of fairness-based retributivism, see, e.g., Sorrell (1999: 18-19) [though see also Murphy (1987) for the claim that Kant, on the whole, is better understood as defending a deterrence view]. And for an insightful critical discussion of the fairness-based position, see Matravers (2000: chap. 2).

and conventions that make such order possible. Society is viewed as the kind of cooperative practice that falls under the scope of the principle of fair play. More specifically, every person who lives within a given legal order enjoys the benefits generated by other people's obedience of the law. It follows from the conjunction of the normative principle of fair play and the descriptive claim on which society falls within that principle's scope that every person who benefits from the legal obedience of others incurs a moral obligation to obey the law herself. And from this it follows that a person who breaks the law acts unfairly. The offender is a free rider on the lawful behavior of others. She derives the same benefits that others derive by enjoying the peace and stability that come from living within the law, but she does not incur the costs that others incur in maintaining the legal order. In doing so, the offender therefore enjoys an unfairly large share of the benefits generated by mutual cooperation.

These extra benefits that the offender unfairly enjoys, it is important to emphasize, should not be identified with the fruits of his offense. A successful thief, of course, does end up enjoying an unfair share of society's economic goods. And a fairness-based retributivist would surely agree that, if caught, the thief should be forced to return the stolen goods for that very reason. But the unfair advantage that the fairness-based retributivist is most concerned with here is that the thief has not accepted the burden of self-restraint assumed by those who obey the law. By violating the law, he has enjoyed a freedom from such restraint that others have not. As Finnis puts the point, the unfair advantage that the offender enjoys is "the advantage of indulging a (wrongful) self-preference, of permitting himself an excessive freedom in choosing," so that he has enjoyed more freedom than has everyone else (1972: 132).[33] The thief takes this unfair benefit for himself regardless of whether, in the end, he benefits economically from his offense. And this unfair benefit, more generally, is enjoyed by everyone who violates the law, regardless of whether their offenses lead them to prosper or to suffer.

From the claim that every offender enjoys an unfair benefit by committing an offense, moreover, it follows that by considerations of fairness, this unfair benefit should be taken away from her. And this, according to the fairness-based retributivist solution, is precisely what punishment accomplishes. When a thief is put in prison, for example, the liberty that she is now deprived of compensates for the extra amount of liberty that she unfairly took for herself in committing her offense. If she is imprisoned for an appropriate amount of time, that is, her imprisonment will restore her to a position in which she has enjoyed just that amount of liberty to which she is entitled by considerations of fairness. At this point,

[33] That this is the benefit at issue is also emphasized by most fairness-based retributivists [see, e.g., Sadurski (1985: 226–7, 229)].

her punishment is completed and justice has been served. As Murphy puts it, "This analysis of punishment regards it as a debt owed to the law-abiding members of one's community; and, once paid, it allows reentry into the community of good citizens on equal status" (1973: 15). Although the fairness-based retributivist may sometimes put this point in the more familiar language of desert, in doing so he means something importantly different from what the more traditional desert-based retributivist means. As Murphy at one point clarifies, the offender "deserves punishment in the sense that he owes payment for the benefits" that he has unfairly enjoyed (1973: 26). Having committed an offense in the past is sufficient to render it morally permissible to punish the affender now, on this account, not because he now deserves to suffer or he has now forfeited some of his former rights, but simply because punishing him now would be fair. Indeed, as Finnis puts the point even more strongly, "it is unfair *ceteris paribus* not to punish the guilty" (1972: 135). It should now be clear what distinguishes fairness-based retributivism from alternative versions of retributivism.[34] And it will also seem clear to many that what distinguishes it from the alternatives also renders it more plausible. The question now is whether it renders it plausible enough.

3.3.2 The Not Punishing the Guilty Objection

The fairness-based retributivist solution turns on two claims: the claim that offenders can be understood as free riders who take unfair advantage of their law-abiding fellow citizens and the claim that if this is so, then punishment is justified as a fair payment that offenders owe non-offenders. Both of these claims may well initially seem plausible. But both, in the end, should be rejected.

The first claim, that offenders can be understood as free riders, is clearly plausible in the context of some legal offenses. Virtually everyone, for example, is likely to feel at least some temptation to cheat, at least a little, on their income taxes. Most people, however, do not give in to this temptation. And it seems plausible to say that a person who does cheat on

[34] Fairness-based retributivism should also be distinguished from the structurally similar but substantively different approach that Lippke has recently proposed as "victim-centered retributivism" (2003b). Both positions attempt to ground a defense of punishment in distributive justice, but while the former appeals to the plausible principle of fair play, the latter depends on the extremely problematic assumption that it is appropriate for the state "to ensure that all citizens are guaranteed equality of condition," at least with respect to certain basic interests (129). This further assumption, when used to justify harm as a means of producing equality, is plainly unacceptable: if a father hits one of his children, is it better that he then hit all the others so that they are made equal in this sense? I will therefore focus my discussion of the distributive justice approach to punishment on the much more promising fairness-based version.

her taxes is taking unfair advantage of everyone else. She allows herself to enjoy a kind of freedom that the rest deny themselves, she benefits from living within the social order that is made possible by their self-restraint, and this accounts for the wrongness of her behavior. Given that all of this seems true, the fairness-based retributivist seems to be in a good position to use this fact to justify the permissibility of punishing the tax evader.

But while characterizing the offense as a form of free riding seems plausible for some legal offenses, it is implausible for many others, including clear cases of some of the most serious and harmful offenses. Consider, for example, what this analysis would imply about rape. Some men rape some women, while many other men rape no women. The claim that legal offenses are to be understood as free riding on the law-abiding would mean that the rapist's offense is to be analyzed in terms of the rapist's unfairly taking advantage of all men who voluntarily refrain from committing rape. But this is unacceptable. As Duff puts it, "this account of the criminal wrongfulness of rape is perverse: what is wrong with rape is that it attacks another person's interests and integrity, not that it takes unfair advantage of the law-abiding" (1986: 212).[35] The same seems to be true of many other serious legal offenses such as murder, assault, and child molestation.

Part of the problem for the fairness-based retributivist here is that the account of legal offenses as a form of free riding seems incapable of capturing what makes these offenses so serious – the great wrongs they do to their victims. But the heart of the problem is that in many such cases, the account depends on a substantial burden to the law-abiding person that simply does not exist. The vast majority of people, for example, have no desire to molest a child. There is therefore no cost to them in refraining from doing so. But if there is no cost to them in refraining, then they bear no burden by refraining, and the offender who does molest a child therefore enjoys no unfair advantage over them.[36] Relatedly, there are some criminal behaviors that many people are incapable of engaging in. Most people, for example, do not know how to hack into the Pentagon's computer system. They are therefore not burdened by a law that forbids them to do so. So, if a hacker succeeds in breaking this law, there is no sense in which they are voluntarily exercising a form of

[35] A number of critics have pointed to the case of rape as undermining the fairness-based retributivist solution in this context, including Duff (1986: 211–17; 1990: 6), Hampton (1988c: 115–16), Walker (1993: 75–6), and Dimock (1997: 58). The same point is also raised by other critics using other examples, such as child molestation and murder [e.g., Narveson (1974: 190), von Hirsch (1993: 7–8), Hershenov (1999), Walker (1999: 599), and Shafer-Landau (2000: 205)]. See also Ezorsky (1972: 367) for a related example that exposes the same basic problem.

[36] This feature of the problem is stressed by Braithwaite and Pettit (1990: 158) and by Cederblom (1995: 309).

self-restraint that the hacker is unfairly refraining from exercising. The fairness-based retributivist defense of punishment is grounded in the claim that breaking the law is a form of free riding. But as these examples clearly demonstrate, there are many serious offenses that do not fit this description. As a result, there are many clear and important cases of lawbreaking for which the fairness-based retributivist cannot justify punishment. And since this is so, the fairness-based retributivist solution will fail the entailment test. Even if this solution can justify punishment in some cases in which it is thought to be deserved, it cannot justify it in all cases that characterize the practice of legal punishment.

3.3.2.1 The General Compliance Response

There are two ways in which the fairness-based retributivist can respond to the problem raised by the claim that many offenders are not free riders. One is to admit that the rapist is not really a free rider in virtue of the fact that other men comply with the law forbidding rape, but to insist that the rapist is a free rider by committing rape nonetheless. On this account, the rapist is a free rider not because the rest of us incur a serious cost in adhering to the law against rape in particular, but rather because we incur a serious cost in adhering to the law in general. Although most of us incur no burden by not committing such offenses as rape, child molestation, and computer hacking, that is, most of us do feel burdened by many other laws that we would rather not obey. We do, however, obey them, and this accounts for the burden that the rapist unfairly refuses to share with us. As Dagger puts the point in raising this defense against the problem posed by the non-free-riding offender, the benefits and burdens that the fairness-based retributivist must appeal to in justifying punishment are not "the benefits provided and burdens imposed by obedience to particular laws" but rather "those that follow from obedience to the laws of a cooperative *practice* – in this case, the rule of law in a reasonably just society" (1993: 481). And on this account, anyone who breaks any law is, for that reason, a free rider. I will refer to this as the "general compliance response."[37]

The general compliance response does, indeed, entail that all offenders are free riders. But it also entails that all offenders are *equally* free riders. And this is why the response must be rejected. The fairness-based retributivist, after all, justifies punishment as a form of fair payment for the offender's having unfairly taken a benefit that the rest of us have not taken. If the unfair benefit taken by a rapist is much greater than the unfair benefit taken by a speeding motorist, then it follows that the fair punishment of a rapist will be much greater than the fair punishment of a speeding motorist. But if, as the general compliance response insists, the

[37] See also Bradley (1999: 108).

unfair benefit taken by the rapist is identical to the unfair benefit taken by the speeding motorist (both have avoided the cost of complying with the law in general, while we have incurred this cost), then it follows that fairness will dictate giving the rapist and the speeding motorist precisely the same punishment. This implication is plainly unacceptable, and so, therefore, is the general compliance response to the not punishing the guilty objection.[38] The response would enable the fairness-based retributivist solution to justify punishing everyone who breaks the law but only by justifying an intuitively inappropriate amount of punishment for too many people. The fairness-based retributivist solution could then pass the entailment test, but at the cost of not passing a reasonable application of the foundational test.

Dagger attempts to respond to this objection to the general compliance response by insisting that even though his position does entail that "the murderer and tax cheater should be punished to the same extent for their crimes of unfairness," it does not follow that "the murderer and the tax cheater should receive the same punishment *tout court*" (1993: 484). For while the murderer and the tax evader are each guilty of a "crime of unfairness," Dagger argues, the murderer is guilty of a second offense: an offense against the person he has murdered. And there is nothing about the fairness-based retributivist position, he insists, that rules out the possibility of aiming a second punishment at the murderer, provided only that it is "justified and established on other grounds" (484). The fact that fairness-based retributivism itself does not justify this further punishment is therefore not a problem. The principle of fairness explains why the state has the right to punish offenders, and that is all that it needs to do. It does not tell us "how exactly to punish wrongdoers," but this is something that it need not do (484).

Dagger is correct in maintaining that a solution to the problem of punishment need not answer every question about how we should punish offenders. But he is mistaken in claiming that this is the only shortcoming that the objection to the general compliance response identifies. For the problem is not simply that the fairness-based retributivist solution fails to provide a further justification for aiming more punishment at the murderer than at the tax evader. The problem is that the principle of fairness necessarily precludes the existence of such a justification. Dagger suggests that the problem is only that the principle of fairness "must be supplemented by other considerations – for example, deterrence, reform, moral education, restitution – when it is time to decide how exactly to punish wrongdoers." However, at the same time, he insists that "none of these other considerations provides a satisfactory account of society's right to punish. For that we must rely on the principle of fair play" (484). But if

[38] This problem has been noted by several writers, including Burgh (1982: 205ff.) and Hampton (1988c: 115).

the fact that punishing a murderer will deter others from committing murder does not justify punishing him in the first place, then the fact that punishing a murderer more than a tax evader will deter others from committing these offenses cannot justify punishing the murderer more than the tax evader. If the only thing that gives the state the right to punish is the state's right to prevent offenders from enjoying an unfair benefit, after all, then the state can punish a particular offender only up to the point at which that offender's unfair benefit has been removed. If, by the general compliance response, the unfair benefit that every offender enjoys is precisely the same, then the state only has the right to punish every offender to the same degree. Punishing the murderer more than the tax evader would therefore be positively unfair. Dagger's defense of the general compliance response, therefore, along with the response itself, must be rejected.

3.3.2.2 The Particular Compliance Response

The general compliance response attempts to show that every offender is a free rider by construing the burden that the offender unfairly refrains from shouldering as the burden of complying with the law in general. In construing the benefits and burdens of legal compliance in this manner, the general compliance response does establish that every offender is a free rider. But in doing so, it also entails that every offender enjoys the same unfair benefit as every other offender and that all offenders therefore merit the same amount of punishment, an implication that the proponent of punishment cannot accept. If the fairness-based retributivist is to overcome the not punishing the guilty objection, therefore, she must establish that a rapist free rides on a burden that nonrapists willingly shoulder, while a speeding motorist free rides on a very different burden that nonspeeding motorists willingly shoulder. And she must explain why the former burden is much greater than the latter. On the face of it, at least, this does not seem possible. The vast majority of nonrapists happily and willingly refrain from committing rape, after all, while the vast majority of nonspeeding motorists obey the speed limit only grudgingly and at considerable inconvenience to themselves. It would therefore seem that, if anything, an analysis of the benefits and burdens of compliance with the law on a case-by-case basis would find that people who break the speed limit or cheat on their income taxes should be punished far more severely than people who commit rape or who molest young children. And this would make the particular compliance response even more objectionable than the general compliance response.[39]

There are two ways in which defenders of the fairness-based retributivist approach have attempted to meet this challenge. Both involve

[39] This point is noted by Dolinko (1991: 545–6).

identifying the benefit that a particular offender unfairly enjoys as the freedom from the particular legal constraint that he violates, but they differ in their analysis of the value of this unfair benefit. One approach is that of Sher, who considers the case of the murderer and the tax evader. Most people find the prohibition on tax evasion more burdensome than the prohibition on murder. They would like very much not to pay their taxes but would not really wish to kill people. So, appealing to the strength of the urge that the offender has allowed himself to indulge will prevent the fairness-based retributivist from accounting for the intuition that the murderer should receive more punishment than the tax evader rather than less. But Sher offers an alternative account of how the fairness-based retributivist can proceed at this point:

> If we believe that the murderer deserves a harsher punishment, it is surely because we regard murder as by far the more seriously wrong act. But if so, then the most natural candidate for what determines the murderer's degree of extra benefit is precisely the strength of the moral prohibition he has violated. By this criterion, the reason he has benefited more is not that he has indulged a stronger inclination, nor yet that he has received greater financial or psychic rewards. It is, instead, that he has violated a moral prohibition of far greater seriousness. (1987: 81)

The true measure of an offender's unfair extra benefit, on this account, lies in the moral wrongness of the offender's unlawful act. The greater the moral wrong his unlawful act does, the greater the benefit he enjoys.[40]

There are two reasons to reject this version of the particular compliance response to the not punishing the guilty objection. The first is that it is entirely ad hoc.[41] For Sher, our belief that the murderer deserves to suffer more than the tax evader is a reason to believe that offenders benefit more from violating morally greater prohibitions. But the first belief can count as a reason for accepting the second belief only if we are already committed to the view that how much an offender deserves to suffer depends on the magnitude of the unfair benefit he enjoys in committing his offense. The fairness-based retributivist clearly cannot assume that this is true, since this would assume that fairness-based retributivism is true. And without this assumption, there is simply no independent reason to believe that one benefits more from violating morally greater prohibitions.[42] Sher attempts to respond to this concern

[40] See also Sadurski (1985: 229) and Sher (1997: 166–7) .

[41] This concern about Sher's position has been raised by a number of writers, including Gert (1989: 427), Braithwaite and Pettit (1990: 159), Murphy (1990c: 282–3), and Zimmerman (1995: 258).

[42] The problem with this version of the particular compliance response seems to be inadvertently acknowledged by Sadurski in his defense of it. In attempting to defend the particular compliance response by appealing to the idea that "criminal law reflects the

by noting that "as the strength of the prohibition increases, so too does the freedom from it which its violation entails," and by concluding that since the murderer "evades a prohibition of far greater force," it follows that "his net gain in freedom remains greater" than that of the tax evader (82).[43] But there are two problems with this further move. First, it is not clear in what sense a murderer enjoys a freedom that a nonmurdering tax evader does not also enjoy. The tax evader, after all, is just as free to commit murder as the murderer is, just as the murderer is just as free to commit tax evasion as the tax evader is. It is simply that while both are free to do both acts, one freely chooses to do the former but not the latter, while the other freely chooses to do the latter but not the former. Indeed, it is not even clear, and for the same reason, why we should think that the murderer enjoys more freedom than the person who commits no offense at all. Even the perfectly law-abiding citizen, after all, is free to commit murder. She simply freely chooses not to do so.[44] Second, even if we agree that there is a sense in which the murderer takes a greater amount of liberty than does the tax evader, this response simply pushes the problem back a step: why is it more *beneficial* to a person to enjoy more liberty rather than less in this sense? If there is no independent reason for people to place more value on enjoying the greater freedom to molest young children than the smaller freedom to cheat on their income taxes, then there will still be no reason to believe that murderers and child molesters gain a greater unfair *benefit* than do speeding motorists and tax evaders, even if the former gain a larger *amount* of freedom than the latter. And if there is no reason to believe this, then there is no plausible

hierarchy of protected values: the more precious the value, the bigger the benefit of non-self-restraint acquired by the criminal," he notes that "[u]nder this conception, it is a mere tautology to say that a more serious crime brings about more benefits of non-self-restraint to the perpetrator" (1985: 229, emphasis added). But that's precisely the problem.

[43] Sher attempts to supplement this argument in a more recent defense of the fairness-based retributivist position by adding the following consideration: "when someone commits murder, he displays not only his willingness to disregard the reasons for rejecting murder as an option, but also, and *a fortiori*, his willingness to disregard the reasons for rejecting acts of any lesser degree of wrongness. He is, in this way, admitting each less wrong act, including shoplifting, into his current option-range" (1997: 173). On this account, too, the murderer takes a greater liberty than does the shoplifter and thus merits greater punishment. But this argument depends on the assumption that everyone who is willing to commit a serious offense would also be willing to commit every less serious offense, and this assumption is unwarranted. A misogynistic rapist might be as unwilling to cheat on his taxes as a tax evader might be unwilling to commit rape [for a similar example, see Ten (2000: 86), who cites the case of a Nazi commander who was responsible for the slaughter of tens of thousands of Jews but who was offended at the suggestion that he or his men would ever stoop to the level of stealing food].

[44] This problem with Sher's position has been noted by Zimmerman (1995: 258).

way for Sher's strategy to give the fairness-based retributivist a satisfactory version of the particular compliance response.[45]

The second problem with Sher's argument is that it only applies to the violation of laws that prohibit morally objectionable behavior.[46] But a law can be just and reasonable even if the act it prohibits is not morally objectionable. There is nothing immoral about driving a car without a license, for example, or building an addition to one's house without a permit. A person who violates a law that requires such legal permissions, therefore, does not enjoy the liberty of violating a moral prohibition. If the benefits that offenders enjoy are construed in terms of the moral wrongness of their behavior, it follows that people who break such laws enjoy no unfair benefit at all and thus cannot be subjected to punishment.

Sher's proposal to measure the value of the liberty that an offender enjoys in terms of the moral wrongness of the offender's act is ultimately unsuccessful. But there is a second way that a defender of the fairness-based retributivist solution can proceed at this point: he can identify the value of a particular form of liberty not with its (im)moral value, but rather with its market value. For any particular illegal act, that is, he can ask how much money people would be willing to pay on an open market for a license to commit that act with impunity and then identify the value of the liberty to commit that act with the fair market price of the license to commit it. The more people would be willing to pay for the freedom to do a particular act, the more the freedom to do that act would be worth, on this account, and the more the freedom to do that act would be worth, the greater the value of the unfair benefit an offender would enjoy. This proposal has been championed by Michael Davis in a series of important papers.[47] Davis imagines a society that determines how much of any particular offense it is willing to tolerate in any given period of time

[45] In the more recent statement of his position, Sher relates the benefit the offender receives more specifically in terms of the benefits of "exercising extra options" that law-abiding citizens refrain from exercising and by claiming that exercising extra options increases people's ability to achieve their goals (1997: 174). But even if all of this is true, it still fails to explain why we should think that a child molester benefits more from exercising the extra option of molesting a child than a tax evader benefits from exercising the extra option of evading her taxes.

[46] Dolinko also makes this point (1991: 546–7). Further difficulties with Sher's version of the fairness-based retributivist position are noted by Ten (1990: 197–200).

[47] Strictly speaking, Davis is not concerned to appeal to fairness as a basis for the state's right to punish, and so, strictly speaking, he is not a fairness-based retributivist. Although he is often cited as such, he has consistently rejected this interpretation of his writings (e.g., 1990a: 226–7). Rather, Davis is concerned with the question of what a fair sentence for a particular offense would be, assuming that punishment is justified. For our purposes, however, we can treat Davis's proposal here as an attempt to help the defender of fairness-based retributivism overcome a problem with his position [see, e.g., Davis (1990a, 1993)].

(perhaps it is willing to bear 100 arsons per month but only 10 murders) and then auctions off one pardon-like license per acceptable incident to the highest bidder. On the assumption that the value of something is "what a willing buyer would pay a willing seller in a relatively fair, efficient, and orderly market," such an auction should provide an "index of the value of the unfair advantage a criminal takes simply by committing a particular crime" (1993: 140).[48]

In one important respect, Davis's proposal represents a significant improvement over Sher's. While it remains obscure at best why we should think that the immorality of a particular act should indicate the value of the liberty to engage in it, it seems quite clear why we should think that the amount of money people would be willing to pay for a particular liberty should indicate its value. Even if we do not think that market value literally constitutes value – that being valuable just is being valued by people – it still seems reasonable to suppose that, at least on average, people would be willing to pay more for more valuable freedoms and less for less valuable freedoms. In this respect, at least, Davis's analysis seems to provide a much more reasonable mechanism for determining what sentence a fairness-based retributivist should endorse for any particular offense. But while the market value approach seems a promising means of generating the sentences justified by fairness-based retributivism, the resulting sentences provide a reason to reject the fairness-based retributivist solution once again.

Consider, for example, the choice between kidnapping (but not harming) a billionaire for a $10 million ransom and torturing a billionaire to death. It seems plausible to suppose that people would be willing to pay more for a license to engage in the former act than the latter. This means that, on the market value approach, the unfair benefit that the kidnapper enjoys turns out to be greater than the unfair benefit that the torturer enjoys. This claim about the relative value of the two freedoms seems plausible enough. The former freedom really does seem to be more valuable than the latter. And so, this might be taken to suggest that the market value approach is on to something. But if the market value approach is accepted as a means of determining the magnitude of an unfair advantage, then under fairness-based retributivism, the punishment for kidnapping (but not harming) a billionaire should be significantly greater than the punishment for torturing him to death. Since this

[48] Davis appeals to this claim in a number of papers, including (1983: 743–5), (1988: 30), (1990b: 239–40), and (1991b: 221). He develops and deploys the auction model in some detail in addressing many legal issues, such as the appropriate amount of punishment for failed attempts, cases involving strict liability, cases involving repeat offenders, and cases involving bad Samaritans in (1986b: 106ff.), (1987: 165ff.), (1985: 133ff.), and (1996c: 104–14), respectively. See also (1986a: 258–60) and (1990b).

implication is plainly unacceptable, so is the market-based version of the particular compliance response to the not punishing the guilty objection.

The problem posed by the case of the kidnapper and the torturer, moreover, is far from isolated.[49] In general, people would be willing to pay more money for licenses to engage in more beneficial acts, and while it is clearly true that in some cases more beneficial offenses are also more serious offenses, there are many other cases in which this is not so. You benefit more from stealing a wallet than from burning down someone's house, for example, and more from overparking than from vandalizing someone's car. The fairness-based retributivist would thus be committed to the view that the punishment should be greater for petty theft than for serious arson and greater for trivial parking offenses than for severe acts of vandalism. And there are, in addition, many other offenses that people commit not because they benefit from them, but simply because they are bored: they damage public property, harass pedestrians, gang up on homeless people, and shoplift products they don't even want, all because they have nothing better to do. These are all significant offenses for which people would be willing to pay virtually nothing for the freedom to engage in and for which the market-based approach therefore cannot justify an intuitively appropriate punishment.

Davis at times seems to believe that he has a satisfactory response to this worry. His response is based on the fact that people can have different motives for bidding on any given license in the kind of auction he envisions. In particular, while some people might want to commit the offense that a given license permits, others might want to keep the license out of the hands of would-be offenders. Consider an extreme version of their problem: child molesting. The vast majority of people have no desire to molest a child. They would pay not even a penny for the opportunity to do so. But a great many people, especially (but not only) those who have young children, strongly desire that children not be molested. A concerned parent participating in Davis's hypothetical auction, then, might very well reason as follows: "If someone like me doesn't purchase the license to molest a child, it will be purchased by someone who will use it to molest a child with impunity. My desire to live in a place where children are safe from being molested is much greater than my desire, say, to speed or to avoid paying my income taxes, and so I will happily pay much more for a license to molest than for a license to speed or to cheat – not because I want to molest a child, but because I want to prevent someone else from doing so. Since there are many other parents

[49] In addition to the cases cited here, see Dimock, who suggests that on Davis's account, the punishment for rape would be less than that for tax evasion for similar reasons (1997: 58). Further useful critiques of Davis's position can be found in Scheid (1990, 1995a, 1995b), Ellis (1997: 84–90), and especially Dolinko (1994).

and concerned citizens who feel as I do, moreover, I will not have to bear this cost alone. We can pool our resources to ensure that we always have enough money to outbid those who would engage in such horrific acts." And so, according to this reasoning, the price for a license to commit such serious offenses as rape, murder, torture, and child molesting will in the end be much higher than the price for licenses to commit less serious offenses like speeding, littering, and tax evasion. The market model for determining the magnitude of unfair advantage can therefore avoid the deeply counterintuitive ranking of offenses that threatens to render it unacceptable.

It is not clear that the considerations invoked by Davis suffice to demonstrate that a license to molest children would cost more than a license to cheat on one's taxes. Without knowing more about the distribution among the bidders of such factors as wealth, aversion to risk, and concern for others, for example, it is difficult to determine whether or not a handful of billionaires might continue to raise their bids on a license to cheat on their taxes long after the asking price had passed the price at which would-be child molesters were forced to call it quits. In addition, without knowing how likely it is that a child molester without a license would be apprehended and how severe the punishment would be, it is difficult to know how much a molester would be willing to pay for a license in the first place. Indeed, without a procedure for determining ahead of time the appropriate punishment for those convicted of any offense without a license, it is not even clear if Davis's proposed model is coherent.[50] Still, even if we assume that Davis is correct to maintain that the presence of people who would bid for licenses to prevent others from using them would suffice to generate an intuitively acceptable ranking of the severity of offenses, this response to the not punishing the guilty objection must still be rejected.

Davis's response ultimately fails because the presence of the bidders it depends on prevents the auction model from fulfilling its intended function: providing a relative value for the various liberties at issue.

[50] For objections along these lines, see Duff (1990: 12–13) and Dolinko (1994: 504ff.). In addition to the problem of determining the level of punishment for those who offend without a license, a further difficulty for Davis's proposal arises from considering what other legal responses, if any, might be appropriate even for those who offend with a license. In (1996c: 110), for example, Davis specifies that committing an offense with a license "pardons only a specific violation of a specific criminal statute" and "leaves all other legal and moral relationship[s] as they were. So, for example, even if a thief 'pays her debt to society' with a 'theft license,' she is still liable to civil suit for return of the property she stole or for damages for any harm she caused." But if a license to steal does not entitle one to keep the stolen goods, then this will render it far less valuable than it would otherwise be, and it becomes far less clear that its market price would be high enough to justify an intuitively appropriate level of punishment for theft.

Rather than ranking the value of the liberty to engage in various offenses, this model ends up ranking something very different: the value of other people not having the liberty to engage in various offenses. As a result, the fact (if it is a fact) that people would pay more for a license to molest children than for a license to cheat on their taxes fails to establish the crucial claim that those who molest children enjoy a more valuable liberty than those who cheat on their taxes. And without that claim, the fairness-based retributivist cannot defend the particular compliance response.

This problem with Davis's solution can perhaps best be seen by first considering a simpler case that Davis himself frequently appeals to: that of hunting licenses (e.g., 1983: 744; 1987: 165; 1990b: 239). So, suppose that a state decides to auction off a specific number of licenses to hunt deer in a given season, and some animal rights groups pool their resources and bid on the licenses to prevent hunters from obtaining them. As a result, the market price for a hunting license goes up from $50 to $100. In this case, it should be clear that the rise in the market price for the license does not reflect a belief of the animal rights groups that the liberty to kill a deer is valuable. They are motivated to raise the price of the license because they believe that it would be better if no one had such a liberty. And if a liberty is such that it would be best if no one had it, then clearly the liberty is not valuable. In the hunting license case, it should be clear that the rise in the market price reflects not the value of people having the liberty, but rather the value of preventing people from having it. This can be seen by considering the fact that if purchasing the right to kill a deer would not prevent anyone else from purchasing this right – if there were no limit on the number of licenses to be sold – hunters would still want to bid on the licenses but the animal rights groups would not. It is only because of the limit on the number of licenses that buying a liberty to kill a deer with impunity also involves buying the power to prevent someone else from acquiring that liberty. And it is only for this reason that the price of a license rises when animal rights groups enter the auction. If we want to know how much people value the liberty to hunt deer, therefore, rather than how much they value the ability to prevent others from enjoying that liberty, we must imagine an auction in which purchasing the former does not automatically entail purchasing the latter.

The liberty to engage in a certain behavior and the ability to prevent someone else from doing so are two fundamentally different things. And this is why Davis's attempt to defend a version of the particular compliance response is ultimately unsuccessful. If I value having a certain liberty, then I incur a significant cost in not having the liberty. If you give yourself the liberty while I decline to enjoy it myself, then you enjoy a significant benefit that I do not enjoy. This is supposed to be the foundation of the fairness-based retributivist position. But if what I value is instead that people not have the liberty, then I do not incur a significant cost in not

having the liberty myself. When someone else gives himself the liberty, then, I cannot complain that he is giving himself a benefit that I have not given myself. When a member of an animal rights group does not obtain the right to go hunting, for example, he does not forgo a liberty that he would be better off having. He does not incur a burden. And so, if a hunter acquires the liberty to go hunting, the activist cannot justly claim that by not giving himself the liberty to hunt, he has incurred a burden that the hunter has not incurred.

But what is true in the simple case of the hunting license is equally true in the more complicated story about how the hypothetical auction for licenses to break the law might yield an intuitively acceptable ranking of the severity of various offenses. In allowing people to bid for a certain liberty not to use it themselves but to prevent someone else from using it, the auction would reveal not that the liberty to molest children is considered more valuable than the liberty to cheat on one's taxes, but rather that the liberty to molest children is not valuable in the first place. But if this is so, then the fact that the license to molest children would (we are assuming) be very expensive will no longer support the claim that child molesters have a more valuable liberty than do tax evaders. A concerned parent who would be willing to spend more money on a molesting license than on a tax-cheating license in Davis's auction would be like a member of the animal rights group who is willing to spend money on a hunting license to deprive a hunter of it. The parent would not place any value on the liberty to molest itself and would think it best if no one had the liberty. Although he would be willing to spend a lot of money for the liberty to molest children in Davis's auction, then, in the world as it is, the concerned parent can justly claim no burden at all in refraining from molesting children, whereas he can legitimately claim to be burdened by not cheating on his income taxes. Since he cannot claim to be burdened by refraining from molesting children, he cannot complain that the child molester is enjoying a liberty that he, the concerned parent, values. And so, in the end, the market-based approach fails to yield an intuitively acceptable version of the specific compliance response. Like Sher's version of that response, Davis's version would enable the fairness-based retributivist solution to pass the entailment test only by preventing it from passing a reasonable application of the foundational test.

While both the general compliance response and the specific compliance response offer ways to justify the claim that the state has the right to punish those who break the law, therefore, neither offers a way to justify this claim without entailing unacceptable claims about the amount of punishment that the state has the right to inflict. And because of this, the problem posed by the non-free-riding offender is, in the end, a sufficient reason to reject the fairness-based retributivist solution. There is no way

to respond to the not punishing the guilty objection that would permit the fairness-based retributivist solution to pass both the foundational test and the entailment test. Thus, there is no way to render it an acceptable solution to the problem of punishment.

3.3.3 The Not Punishing the Previously Victimized Offender Objection

The not punishing the guilty objection rests on the claim that many offenders cannot be characterized as free riders in the sense that fairness-based retributivism requires. I have argued that the fairness-based retributivist cannot provide a satisfactory response to this objection. But there is also another class of offenders that merits attention in this context: offenders who have themselves been the victims of previous offenses. Suppose, for example, that Larry, Moe, and Curly have all been law-abiding citizens, benefiting from the legal order and bearing their fair shares of the burdens involved in maintaining it. Suppose, next, that Larry robs Moe. As an offender, Larry now enjoys an unfairly high share of the costs and benefits involved in social cooperation. As a victim, Moe now enjoys an unfairly low share. And as neither an offender nor a victim, Curly continues to enjoy just the right share. Clearly, the fairness-based retributivist seems to be in a good position to defend the punishment of Larry. But suppose now that Moe decides to commit a legal offense of precisely the same magnitude as the offense that Larry committed against him: he robs Curly. In this case, Moe appropriates for himself an additional amount of personal freedom: more than people are ordinarily allowed to take for themselves, but just enough to make up for the fact that he was victimized by Larry. In this case, the result now seems to be that Moe is back at the level at which he enjoys just the right share of the costs and benefits involved in maintaining the social order. But if he is already at that level, then considerations of fairness can do nothing to justify reducing him to a lower level. And if this is so, then fairness-based retributivism cannot justify punishing him. As Sher puts the problem, "we seem . . . committed to the view that wrongdoers who were themselves previously wronged do not now deserve to be punished" (1987: 85).[51] And since this implication is clearly unacceptable, the case of the previously victimized offender provides a further reason to reject the fairness-based retributivist solution.[52] For many people, the result will be sufficiently

[51] This objection has been endorsed by a number of writers, including Fingarette (1977: 502), Ten (1987: 49), Fouche (1994: 52), and Kershnar (1995: 477–8; 2001: 59–60), and is pressed in particular detail by Anderson (1997).

[52] Sadurski attempts to bite the bullet in response to this objection, conceding that it "may well be a correct implication from the 'balance' model of punishment," at least in the case of burdens that the offender suffers as a result of his offense (e.g., injuries suffered during

counterintuitive to warrant concluding that the solution fails the foundational test. But even those who would be willing to bite the bullet at the level of their moral intuitions would have to concede that this would prevent the solution from passing the entailment test. Punishment, after all, clearly involves the state's having the right to punish those who break the law, regardless of whether those lawbreakers have been previously victimized by other lawbreakers.

Sher responds to this objection to fairness-based retributivism by arguing as follows:

> Even if *X* has previously wronged *Y*, it hardly follows that a fair balance of benefits and burdens is restored when *Y* in turn wrongs *Z*. If *Y* does this, then the original wrongdoer *X* is still left with the double benefit of moral restraint upon others plus his own freedom from such restraint; and the current victim *Z* is left with the double burden of moral restraint on his acts plus the absence of restraint on the acts of (some) others. Thus, the original unfairness is not removed but merely displaced. (85)

But this response is inadequate. It establishes that if Moe commits an offense against Curly, this will not make things *better* with respect to fairness than if he does not. The response therefore shows that fairness-based retributivism does not entail that Moe *should* commit an offense against Curly. But the objection to fairness-based retributivism arising from an offender's previously being wronged is not that it entails that Moe should now commit an offense against Curly. The objection is that it entails that if Moe does commit an offense against Curly, then Moe may not be punished. The objection, that is, is not that Moe's wronging Curly will increase fairness; it is simply that it will not make things any worse. If Moe's committing an offense against Curly does not reduce fairness, after all, then considerations of fairness will provide no grounds for punishing Moe to make up for it. And the fact that Moe's wronging Curly does not increase fairness does nothing to support the claim that it reduces fairness. Indeed, Sher himself concedes that, at most, we can say that the unfairness is merely *displaced* by the second offender's act. This much seems to be correct. But this then seems to entail that from the point of view of fairness, we must be indifferent between the state of affairs in which Larry offends against Moe, on the one hand, and the state of affairs in which Larry offends against Moe and Moe then offends against Curly, on the other. In each case, Larry enjoys an unfair benefit

his escape, apprehension, interrogation) but insisting that if it is so, then judges and juries should simply take this into account in sentencing: whatever burden the offender has already suffered "reduces an overall amount of the benefits he has acquired" (1985: 230–1). But on this account, there will be too many cases in which the offender has received no overall benefit and the state thus has no right to punish at all.

and the last person wronged suffers an unfair burden (Moe in the first case, Curly in the second). So, in each case, fairness provides a potential justification for punishing Larry. But in the second case, Moe enjoys exactly the amount of liberty to which fairness entitles him, so considerations of fairness provide no support for punishing him for wronging Curly. The result – that the state may not punish Moe for robbing Curly – is unacceptable. For most people, it violates their moral intuitions and, as a means of justifying punishment, it draws the line between those the state may punish and those it may not in the wrong place. Thus, even those who resist conceding that the fairness-based solution fails the foundational test still have to acknowledge that it fails the entailment test. And, either way, the problem of the previously victimized offender provides a second reason to reject the fairness-based retributivist solution to the problem of punishment.[53]

3.3.4 The No Excuses Objection

I have argued that the analysis of legal offenses as a form of free riding must be rejected for many serious offenses such as rape and murder, and that it must also be rejected for offenders who themselves have been victims of comparable offenses. A further problem concerns offenses committed under what would typically be regarded as mitigating circumstances. Consider, for example, the law against physical assault. The law forbids me from physically attacking you if you have done nothing to provoke me, but it also forbids me from physically attacking you if you have provoked me. I do not feel particularly burdened by the prohibition on unprovoked attacks. So, a person who attacks another without provocation is not unfairly enjoying a liberty that I would like to have for

[53] Sher's response to this objection has also been criticized by Garcia (1989: 274) and Kershnar (1995: 477–8; 1997a: 75–91). In a more recent response to the not punishing the previously victimized offender objection, Sher maintains that the objection can be overcome by appealing to the claim that the relevant principle of fairness applies to pairs of individuals rather than to groups as a whole (1997: 176). On this account, he argues, when Larry robs Moe and Moe then robs Curly, the fact that Moe was robbed by Larry does nothing to eliminate the unfairness within the Moe–Curly relation generated by Moe's robbing of Curly. But this response leaves the fairness-based retributivist open to a narrower version of the objection: the position cannot justify punishment if the offender was previously victimized by the victim. Suppose, for example, that Larry burns down Moe's house. Two weeks later, Moe burns down Larry's house. When Moe burns down Larry's house, he clearly breaks the law. If the state has the right to punish people for breaking the law, then it clearly has the right to punish Moe for burning down Larry's house. But on Sher's more recent "fairness between pairs of individuals" version of the fairness-based retributivist position, there is no unfair balance in the relation between Larry and Moe after Moe burns down Larry's house, and thus no basis for the claim that the state has a right to punish Larry.

myself. But a person who attacks another because he was seriously provoked does help himself to a liberty that I and many other people would like to have. This suggests that the unfair benefit that a provoked attacker enjoys is greater than the unfair benefit that the unprovoked attacker enjoys. This, in turn, means that under the fairness-based retributivist solution, a provoked attacker should receive more punishment than an unprovoked attacker. And since this result seems clearly unacceptable, the problem of excusing conditions provides a further reason to think that the solution cannot pass the foundational test.

The problem for the fairness-based retributivist becomes even greater in circumstances commonly viewed as fully excusing rather than simply mitigating responsibility. Suppose, for example, that Larry becomes temporarily insane and kills Moe. Commonsense morality and law as it is generally practiced dictate that if Larry was insane at the time of his action, he should not be punished. But whether or not Larry was insane, the fact remains that he enjoyed a great liberty that the rest of us do not. To allow him to go unpunished would therefore be unfair to the rest of us. And so, on the fairness-based retributivist account, it follows that he should be punished and that, more generally, people should be punished for their offenses even when they are not responsible for them. This provides a further reason to reject the analysis of legal offenses in terms of fairness.[54]

3.3.5 The Free Rider's Rights Objection

The fairness-based retributivist solution is grounded in the claim that offenders are best understood as free riders. I have argued that there are many reasons for concluding that this claim is false. But let us now suppose that the claim is true: every person who breaks the law does indeed free ride on those who do not; the more serious the offense, the greater

[54] Sadurski attempts to respond to this objection by maintaining that an "insane person, or a person acting under duress, does not *act* in any sense which might be regarded as enjoyment of his freedom. The benefits which are acquired by a criminal are the benefits of unrestricted liberty, yet this state of affairs does not occur in the case of an insane killer, or a person acting under coercion. In their cases there is no broadening of their freedom. In no way were the constraints on their actions reduced" (1985: 230). But this response seems arbitrary. A provoked person who attacks his provoker under duress enjoys a liberty that provoked people who refrain in the same situation do not. An insane person who kills people enjoys a liberty that other insane people who refrain from killing (say, out of an irrational fear that aliens will abduct them if they do) do not. If being open to more options rather than fewer ones is always a benefit to the individual, as the fairness-based retributivist must insist to avoid the problem posed by the non-free-riding offender objection, then there is no reason to think that people who break the law for what are typically taken to be partially or fully excusing reasons do not also enjoy such unfair benefits.

the unfair advantage they enjoy; and those who have excuses benefit less from their offenses than those who do not. Even if all of this is true, the argument for fairness-based retributivism must still be rejected. For it does not follow from one person's free riding on the sacrifices of others that it is morally permissible to coercively extract payment from her for doing so.[55] And yet punishment, on the fairness-based account, is precisely such a form of coercively extracted payment and thus requires precisely this inference. Since the inference itself is unwarranted, it follows that even if the foundation of the fairness-based retributivist solution can be rendered sufficiently plausible, it cannot be used as a basis for justifying punishment. The solution will therefore fail the entailment test even if it passes the foundational test and will have to be rejected for that reason.

To see this further problem with the fairness-based retributivist approach more clearly, an example may be helpful. Consider the following scenario. You live in an apartment building on a dangerous block in a dangerous neighborhood. Every week for the past year, someone on your block has been victimized by a serious offense: robbery, assault, vandalism, and so on. You carry a considerable burden in terms of personal insecurity by living on this block, but you cannot afford to move to a safer neighborhood. One day, a neighbor organizes a block patrol. Every day and every night, for four hours, two people who live on the block volunteer to patrol the street and the hallways in the buildings. The crime rate on the block plummets. You receive a great benefit from the drastic reduction of crime. And this benefit has been made possible only by the fact that most people living on the block have incurred a cost by giving some of their time to serve on the patrols. It follows from all of this that if you decide not to participate in the patrol, you will be a free rider, unfairly benefiting from the sacrifice of others without making a comparable sacrifice of your own. Let us suppose that this is true. Still, it clearly does not follow that it would be morally permissible for your neighbors to coerce you into joining the patrol or to coercively extract payment from you to render the distribution of benefits and burdens on the block more fair (say, by taking enough money from you to pay someone else to take your place). You are being unfair to your neighbors if you do not do your fair share to keep the patrol going, since you benefit considerably from it, and your neighbors have the right to criticize you as a free rider for that reason. But they do not have the right to do more than that. And the same must therefore be said about those who break the law on the fairness-based retributivist account. Even if we accept the claim that everyone who breaks the law is a free rider, we must conclude that even though we could make the distribution of benefits and burdens in society more fair by punishing them, this does not make it permissible for us to do so.

[55] This problem has been noted by Ellis (1997: 93–5).

3.3.6 The Punishing the Innocent Objection

Let us now suppose that the objection grounded in the rights of free riders can also be overcome and that considerations of fairness do justify punishing people who break the law. A further problem will then arise. For if this is so, then it will once again be permissible to punish some people who do not break the law as well. There are, in fact, two different cases in which fairness-based retributivism would justify punishing innocent people, causing it to fail the entailment test and, for most people, the foundational test as well.

One scenario was presented earlier in the context of desert-based retributivism. In this case, the state can punish one innocent person to apprehend and punish five guilty people. If punishment is justified by considerations of fairness, then it would seem that the state may go ahead and punish the innocent person in such cases. If the state does punish the innocent person, then one person will have an unfair balance of benefits and burdens, but if it does not, then five other people will have this unfair balance. If the state has the right to harm people to bring about a fairer distribution of benefits and burdens, then it has the right to punish the innocent in cases such as this, because in these cases the overall distribution of benefits and burdens will be fairer if the innocent person is punished than if he is not.[56]

The second scenario is more specific to the fairness-based retributivist solution in particular. If the state has the right to punish offenders because they are free riders, then it has the right to punish nonoffenders who are free riders for the same reason. And the same argument that establishes that breaking the law is unfair also establishes that many forms of legal behavior are also unfair. Consider, for example, promise keeping. People benefit in innumerable ways from it. This practice is sustained only because most people most of the time keep their promises. And yet, keeping promises is often burdensome. It follows that we benefit from the fact that other people typically incur the burdens of keeping their promises. A person who breaks her promise, therefore, unfairly enjoys a liberty that other people do not. She is a free rider in just the same sense, and for just the same reason, as the person who breaks the law. If an offender's status as a free rider is sufficient to justify the permissibility of the state's harming him to deprive him of his unfair benefit, however, then the same justification applies in the case of the law-abiding promise breaker, another free rider with an unfair benefit.[57] This again amounts to punishing innocent people, and as we have already seen, any defense

[56] Braithwaite and Pettit (1990: 49, 51) make this point.

[57] The same would hold for the person who refuses to participate in the neighborhood patrol program.

of punishment that entails this is unacceptable for that reason. At the very least, it means that the fairness-based retributivist solution fails the entailment test. In addition, for most people, it means that it fails the foundational test as well.

3.3.7 The Harm Versus Punishment Objection

I have argued that legal offenses should not be understood in terms of an offender's taking unfair advantage of the law-abiding: that even if they should, this would still not justify coercively depriving offenders of their unfair shares of the benefits of social cooperation; and that even if it did, it would objectionably entail that we should also sometimes deprive legally innocent people of part of their share. But let us now, finally, suppose that I have been mistaken about every objection that I have raised against the fairness-based retributivist position: every (and only every) offender does enjoy an unfair benefit, this justifies coercively depriving him of that benefit, and doing so does not entail any objectionable treatment of nonoffenders. Even if all of this is true, the fairness-based version of retributivism will still fail to solve the problem of punishment. It will, if we grant all these assumptions, justify the claim that we are entitled to deprive offenders of the benefits they unfairly enjoy. And in doing so, it will justify the claim that we are entitled to act in ways that will harm offenders. But this will still fall short of justifying the claim that we are entitled to punish them. And because of this it will, one final time, fail the entailment test.

The reason for this is the same as the final reason given for rejecting the forfeiture-based retributivist solution: there is a difference between harming offenders and intentionally harming them. The argument that it is permissible for the state to do an act that will predictably harm a citizen does not establish that it is permissible for the state to do an act that will intentionally harm her. Only a justification of intentionally harming people for breaking the law will count as a justification of punishment. But even if fairness-based retributivism can justify harming offenders, it cannot justify intentionally harming them.[58] And so, even if it can justify

[58] Defenders of fairness-based retributivism have at times come very close to acknowledging that this is a feature of their position without recognizing that it is, at the same time, a problem with it. In his defense of fairness-based retributivism, for example, Sher cites Morris's article approvingly: "[Morris's] suggestion immediately moves the discussion to a new level. By de-coupling punishment from pain and suffering, it defuses the charge that all attempts to exact retribution are mere senseless cruelty" (1987: 77). Sher's idea here seems to be that the pain and suffering involved on the fairness-based retributivist account are not really aimed at, but are simply the inevitable by-product of taking away the criminal's unfair excess benefit. This does seem to be a correct characterization of the proposal and it does, indeed, seem to make the proposal more humane than

doing acts that would harm all and only offenders, it is still not a solution to the problem of punishment.

To see that this is so, it may help to consider a case in which the claim that one person is enjoying an unfair share of the benefits of a particular form of social cooperation is uncontroversial. So, suppose that four friends pool their money to purchase one set of season tickets to the local baseball games. Since the best seats are reserved for season ticket holders, and since the price per game is considerably lower if they buy tickets for the whole season rather than for single games, each person in the group benefits significantly from the contribution of the others. Suppose, for example, that the team will play 100 home games during the season, a season ticket for one good seat for each game costs $2,000, and individual tickets for a comparable seat for a single game cost $30 each. In that case, it would cost each person $750 to buy tickets to see twenty-five games on his own, but would cost each person only $500 to buy tickets for comparable seats to see twenty-five games as part of a season ticket. In this case, each of the four friends clearly benefits from participating in this particular form of mutual cooperation, and it is easy to determine the fair share that each should have to incur to enjoy this benefit: since the season ticket costs $2,000 and since they each enjoy an equal share of the benefit produced by their joint purchase, fairness dictates that each incur a $500 cost in order to enjoy the benefit.

Now, suppose that one of the four friends has refused to pay his fair share. Larry put the order on his credit card; Moe and Curly each wrote him a check for $500, but Shemp has thus far contributed only $150 even though he has received a full share of the tickets. Larry and Shemp are unable to resolve their disagreement, so they go to a mediator and agree to be bound by the mediator's decision. Given the facts set out here, the mediator determines that to be fair, Shemp must give Larry another $350. Since Shemp has agreed to be bound by the mediator's decision, this means that the mediator is forcing Shemp to pay Larry $350. But while the mediator's act harms Shemp, and while the mediator understands this, the mediator does not act with the intent of harming Shemp. He orders Shemp to hand over the money because Shemp is currently

retributivism has often seemed to be. But it accomplishes this precisely because it makes the proposal less punitive than retributivism requires it to be. Similarly, Murphy characterizes the fairness-based approach as one on which the offender "deserves" punishment only "in the sense that he owes payment for the benefits" he has unfairly enjoyed (1973: 26). And Davis characterizes the unfair advantage account of sentencing as one on which the criminal law operates as "a system of administered prices. . . . A penalty is a fair price only if it corresponds to what a license to do that crime would fetch on the open market" (1983: 745). This can again make the fairness-based approach seem more reasonable than some of its more familiar rivals, but again, only because the harm that the offender suffers is not aimed at, but is simply the unavoidable result of forcing him to pay a fair price for the benefits he has enjoyed.

holding on to more than his fair share of it, and in doing so, the mediator foresees that this will be harmful to Shemp. But this is clearly not a case in which Shemp is punished.

And now, finally, consider that, on the fairness-based retributivist position, what the mediator does to Shemp when he orders him to pay Larry the $350 is precisely what a judge would do to an offender when she ordered him to serve a sentence. Suppose, for example, that it makes sense to say that an arsonist, by committing arson, has enjoyed an unfair excess amount of freedom that is worth five years of freedom more than nonarsonists enjoy. And suppose, in addition, that because he enjoys this unfair excess freedom, a judge is now entitled to put him in prison for five years. If this is so, on the fairness-based retributivist account, then the judge who takes away the arsonist's freedom does so with the understanding that this will harm the arsonist, but not with the intention of harming him. The judge does not, on the fairness-based retributivist account, take away the arsonist's freedom in order to make him suffer; she does so only to ensure that the arsonist does not keep more than his fair share of the benefits generated by the legal order. And so, in this case, just as in the case of the season ticket, the arsonist is not punished. He is harmed, certainly, but he is not punished. And thus, even if the fairness-based retributivist solution can overcome all of the other objections that have been raised against it in this section, it is still not a satisfactory solution to the problem of punishment.

3.4 OTHER VERSIONS OF RETRIBUTIVISM

I have argued against the retributivist solution to the problem of punishment by challenging the three most prominent versions of retributivism: those that appeal to considerations of desert, rights, and fairness. Before discussing those defenses of punishment that are best understood as neither consequentialist nor retributivist, however, I will conclude this chapter by briefly considering a few other versions of the retributivist solution.

3.4.1 Trust-Based Retributivism[59]

Another version of retributivism is grounded in considerations about the nature and value of trust. Trust between people is essential to the proper functioning of a civil society. We trust others to behave appropriately

[59] The discussion in this section is deeply indebted to Dan Korman's excellent paper "The Failure of Trust-Based Retributivism" (2003), which demonstrates how some of the objections I raised in previous sections of this chapter can be aimed at this further form of retributivism. For further criticisms of the trust-based retributivist solution, see also Gavison (1991: 355ff.). And for a defense of the claim that considerations about promoting trust actually justify the abolition of punishment, see Fatic (1995, esp. chaps. 8 and 9).

when we drive, when we do business, when we walk through the park, even when we enjoy the privacy of our own homes. As Bok has put it, "Whatever matters to human beings, trust is the atmosphere in which it thrives" (1978: 31n). But trust can be sustained only to the extent that people behave in trustworthy ways, which people who break the law fail to do. On this trust-based version of the retributivist position, then, what distinguishes offenders from nonoffenders is that offenders violate the trust of others, and punishment is justified as the appropriate response to such violation. As one proponent of this trust-based solution to the problem of punishment has put it, "what makes an act an appropriate ground for punishment is *the betrayal of trust*" [Hoekema (1991: 345)].[60]

3.4.1.1 The Punishing the Innocent Objection

The position that betraying trust is sufficient to render one liable to punishment is, in its unqualified form, clearly subject to the punishing the innocent objection. People who cheat on their spouses, for example, or who are chronically late for appointments betray the trust of others and thereby undermine the intricate web of social relations that makes trust possible. Unless it can somehow be modified, therefore, trust-based retributivism will unacceptably entail that it is permissible for the state to punish these people even though they have broken no laws. And this, in turn, would mean that the trust-based retributivist solution fails the entailment test and, for most people, the foundational test as well. The question, then, is whether the trust-based retributivist can provide a satisfactory account of the distinction between trust violations that render punishment permissible and those that do not.

One possible response to this problem is to appeal to the amount of harm caused by a given act of trust betrayal. This response is clearly unsatisfactory. Losing twenty dollers to a pickpocket is less harmful than being betrayed by a lover, but it is the punishment of the pickpocket, and not of the unfaithful lover, that the trust-based retributivist must account for. A second possibility is to appeal to the distinction between voluntary and involuntary trust relationships. When my friend promises to help me paint my house next weekend, I am free to decide whether or not to trust that he will show up, but when a pharmacist tells me that the pills in the bottle are the medicine I have ordered, I have no choice but to trust her. If the legal system is understood to be concerned only with the preservation of unavoidable trust relationships, then this distinction can be used to account for why punishment would be inappropriate if my friend lets me down but appropriate if my pharmacist does. This response has been

[60] Similarly, Dimock maintains that "the violation of conditions of basic trust in a community is the characteristic wrong which criminal behaviour involves, and it is this upon which the justification of punishment rests" (1997: 39).

defended by Hoekema as follows: "The reason that punishment is inappropriate [in some cases of trust violation but not in others] has to do not with the gravity of the harm caused, but with the voluntary character of the trust relationship" (1991: 347).

The problem with appealing to the distinction between voluntary and involuntary trust, however, is that it, too, fails to draw the line in the place where the defender of punishment seeks to draw it. A woman who wishes to live in society, for example, has no choice but to trust that the men she shares the streets and sidewalks with will not rape her. But a woman can choose whether or not to trust her boyfriend enough to invite him to her apartment alone at night. This shows that in Hoekema's version of the trust-based retributivist position, the state has the right to punish a stranger for raping a woman in public but not for raping a girlfriend who trusted him enough to let him into her apartment. And this is clearly an intolerable implication.[61]

A final response to the problem of demarcating punishable from nonpunishable trust violations has been defended by Dimock (1997). On Dimock's account, the relevant distinction is not between more and less harmful trust violations or between voluntary and involuntary trust relationships, but rather between what she calls the "subjective" and "objective" conditions of trust (45ff.). The subjective conditions of trust are those that "actually cause individuals to trust others," while the objective conditions of trust are those "that serve to make trust in others objectively reasonable" (1997: 46, 51). One person might trust another even though there was an objectively good reason to distrust him, for example, and one might distrust another even though there was an objectively good reason to trust him. This distinction helps the trust-based retributivist to solve the demarcation problem, Dimock argues, because the law can be understood as "an institution that makes trust *objectively* more reasonable" (1997: 51, emphasis added). And understanding it in this way, she maintains, provides a satisfactory answer to "the question of which violations of trust merit the attention of the law" (1997: 50).

Dimock attempts to sustain this position by appealing to a set of cases. A wife who commits adultery violates the trust of her husband, for example, as does a wife who deliberately infects her husband with HIV. A person violates the trust of a stranger by deliberately giving him false

[61] Relatedly, as Korman has pointed out (2003: 564–6), Hoekema's distinction fails even in his own case involving the pharmacist. A professional chemist might be able to verify that the pharmacist had correctly filled her order and choose to trust the pharmacist because this would be more convenient than testing the pills herself. This entails that Hoekema's version of the trust-based retributivist solution would justify punishing a pharmacist who defrauded a customer who lacked the ability to verify the order, but it could not explain why the state would have the right to punish a pharmacist who defrauded a professional chemist.

information when he is asked for directions, as does a person who robs a stranger. For Dimock, it is clear that a successful defense of legal punishment will have to explain why punishment is permissible in the latter member of each pair of cases but not in the former.[62] And she claims that the appeal to the relevance of the objective conditions of trust provides the explanation. For in the case of the husband who is infected by his wife and the person who is robbed by a stranger, she argues, "if society acquiesced in their being so treated, this would make everyone less trustworthy in an objective sense. Knowing that our fellows have allowed us to be victimized without complaint and protest and condemnation, or that they are unwilling to assist us in providing protection against further abuse, would make trust of them, and not just of our violator, less *objectively* reasonable" (1997: 51, emphasis added). And since, for Dimock, "the law ought to be concerned" with those violations of trust "that make mistrust more objectively reasonable than it would otherwise be" (51), it follows, on her account, that in these cases the law has the right to punish the offenders.

The problem with Dimock's argument, however, is that what she says about people who break the law is equally true of many people who do not. Indeed, it is true of the very cases that Dimock herself cites as cases of innocent people who should not be punished.[63] If nobody complains about a woman who cheats on her husband or about a man who deliberately lies to a stranger, then it is objectively more reasonable to believe that the adulterer and the liar will continue in their ways, and that others will as well. In a world where people were punished every time they told a lie or cheated on a lover, it would be objectively more reasonable to trust that people would tell the truth and be faithful to their partners. And so, in the end, the distinction between objective and subjective conditions of trust fares no better than the alternatives in fending off the punishing the innocent objection.[64] The line between those who undermine trust and those who do not simply cannot be drawn where the practice of legal punishment requires.

[62] I will therefore assume that there is a (just and reasonable) law prohibiting what the wife and stranger do in the latter case but not in the former, though some may find the example itself contentious.

[63] This problem is also noted by Korman (2003: 567–8).

[64] It is also worth noting in this context that even though Dimock explicitly defends her trust-based justification of punishment in retributivist terms, it seems also to be subject to the version of the punishing the innocent objection that undermines consequentialist solutions to the problem of punishment: if deliberately framing and punishing one innocent (i.e., non-trust-violating) person would help to deter many other people from violating the trust of others, then punishing that innocent person would make it more objectively reasonable for people to trust others.

3.4.1.2 The Not Punishing the Guilty Objection

I have argued that trust-based retributivism is subject to the punishing the innocent objection and that the various attempts to overcome this objection have been unsuccessful. Like many of the other solutions subject to this objection, the trust-based retributivist solution also incurs the problem that for some people who break the law, punishment is not justified. Trust-based retributivism is based on the claim that "the person who violates the criminal law, in the absence of excusing conditions, demonstrates conclusively that there are reasons not to trust her" [Dimock (1997: 53)]. But there can clearly be offenders who are perfectly trustworthy for the simple reason that they promise ahead of time to commit their offenses and then do exactly that. Consider, for example, a physically intimidating thug who runs an extortion racket. He approaches all of the shop owners in town and makes each of them the following deal: pay me $100 each month, and I will make sure that you don't have any trouble; fail to pay, and I will make sure that you do. The extortionist is successful precisely because he is trustworthy: everyone who pays gets protection, and everyone who doesn't gets trouble. Since the extortionist does only what he says he will do, his behavior does nothing to diminish the objective grounds that people in his community have for trusting one another (indeed, if anything, his perfectly predictable behavior *increases* the objective reasons that people have for trusting others), and the same can be said for many other cases: the boyfriend who promises to be nice to his girlfriend as long as she doesn't talk to other men and to beat her if she does, the racist who promises to leave blacks alone as long as they stay away from his neighborhood but to shoot them if they don't, and so on. Since these are clearly cases in which a defender of punishment must justify punishing people for their unlawful behavior, but since they are also cases in which the people in question do nothing to undermine the objective conditions of trust, they provide a second reason for concluding that the trust-based retributivist solution fails the entailment test and fails a reasonable application of the foundational test as well.

3.4.1.3 The Harm versus Punishment Objection

I have argued that trust-based retributivism justifies harming some nonoffenders and fails to justify harming some offenders. But let us now suppose that I have been mistaken about both of these claims, and that the line between trust violators and nontrust violators can be justifiably drawn in a way that perfectly maps onto the line between offenders and nonoffenders. Even if this is so, there is still one final reason to reject the trust-based retributivist solution. The reason is that even if all of this is so, it will justify harming offenders but it will not justify punishing them.[65]

[65] This problem is also noted by Korman (2003: 570–5).

The reason for this becomes clear when one looks carefully at the move that trust-based retributivists attempt to make from the claim that an offender has violated trust to the conclusion that it is permissible (or even obligatory) to punish her. Dimock, for example, says that "if society acquiesced" in a particular violation of trust, if society allowed the trust violation to occur "without complaint and protest and condemnation" (1997: 51), and if "such behaviour is tolerated" (52), then the objective conditions of trust would break down. But there is a large gap between the claim that if society did nothing in the face of lawbreaking then trust would be undermined and the claim that if society did not respond to such offenses with punishment, in particular, then trust would be undermined. The trust-based retributivist thesis requires a defense of the latter claim, but its supporters have at most offered a defense of the former.[66] And the move from the former to the latter is unwarranted.

Consider, for example, that, in order to park my car and leave it on the street, I must trust that others will refrain from vandalizing it while I am gone. If the state responds to vandalism by simply acquiescing and tolerating it, then I will have little reason to trust people to leave my car alone. But if, for example, the state responds to vandalism by compelling offenders to repair all of the damage they have caused, by taking active measures to prevent vandalism, by educating people not to vandalize, and so on, then there is no reason to suppose that this will fail to establish the trust that writers such as Dimock and Hoekema wish to preserve. Punishment, as we have seen at various points in this book, involves not merely performing an act that harms an offender but doing so with the intention of harming an offender. And while considerations of trust might justify the former, this is still not sufficient to justify the latter.

Indeed, Dimock herself comes very close to recognizing this crucial limitation on her position when she finally considers the forms of treatment that might be justified by her position in particular cases. For such financial offenses as embezzlement and tax evasion, for example, she asks, "What would be necessary to reestablish objective trust in the face of such violators?" and suggests such possibilities as barring an offender from holding a certain position or from operating a certain business (55–6). Even for such offenses as burglary and robbery, she suggests that her position would often justify something other than incarceration: "Employment and educational programmes under suitable supervision seem more likely to facilitate the goal of reestablishing trust in communities when dealing with such

[66] Hoekema's explanation of why punishment, in particular, is warranted suffers from the same defect (1991: 348–5): he argues that we must do something to indicate that trust has been betrayed, to prevent offenders from doing so again, and to encourage people to trust one another, but he does not explain why punishment in particular is justified on such grounds.

offenders" (56). But giving a thief a job in order to reestablish trust in the community is not the same as punishing her. Depriving a restaurant owner of a liquor license is an act of punishment if it is done to harm him, but not if it is done to restore trust in the community and the harm is merely a foreseen consequence. Finally, even in cases where Dimock thinks that incarceration may be warranted, it is still justified only so that certain offenders can "be removed from the community" (57). In these cases, too, the harm to the offender will be merely foreseen rather than intended (this, after all, is also why we quarantine people), and so will again fail to count as punishment. For this reason, then, trust-based retributivism fails the entailment test once more, and so again cannot provide a satisfactory solution to the problem of punishment.

3.4.2 Debt-Based Retributivism

A further version of retributivism has more recently been defended by McDermott (2001). On McDermott's account, when an offender violates a victim's rights, the victim suffers two kinds of loss: a material loss and a moral loss (411, 415). If, to follow one of McDermott's examples, Barry steals something that belongs to Arthur, then Arthur "has suffered two losses. First, he has lost [the object]; second, he has not received the treatment due to him as a right-holder" (411). Each of these things is something that Arthur has a right to, and so each, in effect, is something that Barry must either give back or replace. As McDermott puts it, "Since Arthur is entitled to both of these goods, Barry incurs debts for the value of both of these goods – a debt for the value of X, *and* a debt for the value of the treatment he withheld from Arthur by violating his right" (411).

McDermott acknowledges that an offender could fully discharge the first debt without undergoing punishment. If the object in question is worth $1,000, for example, then Barry could fully discharge his material debt to Arthur by giving him $1,000. But, McDermott argues, the second debt, the moral debt, cannot be discharged in a similar manner. "Money may be valuable," he notes, "but it is not valuable in the *same way* as the treatment the wrongdoer withheld from his victim, and therefore transferring money to the victim will do nothing to settle the wrongdoer's debt" (414). This is so, on McDermott's account, because moral goods "are valuable for the very reason that they *distinguish* the recipients as members of our moral community. The value of being treated as a person, in other words, is that such treatment distinguishes one as a person." Offenders, therefore, "cannot settle their moral debts by unilaterally transferring material goods ... because providing others with material goods does *nothing* to distinguish them as members of our moral community, and therefore transferring a material good will do *nothing* to settle a moral debt" (418).

As a result of this part of McDermott's argument, the offender owes something of value to the victim and cannot discharge this debt simply by giving the victim material goods. Barry has a debt to Arthur; its value is the value of the moral good that Barry denied to Arthur, and the value of this moral good is different from the value of material goods. Now the moral good that Barry deprived Arthur of, the good of being treated as a person, as a right holder, is one that Barry himself has. So, if Barry could "transfer" his moral good to Arthur, then he could pay back his second debt as easily as the first. But, McDermott states, moral goods are not transferable (419). The result, McDermott concludes, is that Barry is currently enjoying a moral good that he would be obligated to transfer to Arthur if this were possible. But it is not possible. In relevantly similar circumstances, we would not allow Barry to keep something that he would have to transfer to Arthur if this were possible. If Barry stole something from Arthur and Arthur died before Barry could be made to return it to him, for example, Barry would then be enjoying the use of a good that could not be transferred to Arthur but that he would have to transfer to Arthur if this were possible [i.e., if Arthur were still alive (422)]. In this case, McDermott says, it should be clear that even though Barry cannot return the good to Arthur, Barry has no moral right to keep it. The debt that Barry owes to Arthur, unpayable as it is, still means that Barry has lost the right to the good even if he cannot transfer it to Arthur. And so, McDermott concludes, the same must be said of the moral debt that Barry cannot pay in the original case, in which he cannot return the good because moral goods are by their very nature nontransferable: Barry is now enjoying a moral good that he would have to give to Arthur if he could, and the fact that he can't doesn't mean that he has any moral right to keep it. It is morally permissible to take that moral good from him. This would amount to treating him in a manner that would ordinarily count as a rights violation. And this, on McDermott's account, is precisely how punishment is ultimately justified on retributivist grounds: punishment, that is, "is a means of denying these forfeited moral goods to the wrongdoers" (424).

This debt-based version of retributivism is interesting, and in some respects it seems to avoid certain problems that undermine some of the more familiar versions of retributivism. In the end, however, it is unsatisfactory. The first problem with the debt-based retributivist solution is that it fails to establish that moral debts cannot be repaid. McDermott argues that moral goods are valuable because they distinguish us as members of the moral community, and that merely transferring material goods to a person cannot do this. But this establishes only that an offender cannot repay his moral debt by writing a check. It does not establish that he cannot repay it at all. If part of the reason Arthur is worse off than he was before Barry robbed him is that Arthur is now less fully

affirmed as a moral agent, then this can generate in Barry a debt to take other actions that would reaffirm this. He might be obligated to apologize to Arthur, to publicly affirm Arthur's moral standing, to promote some of Arthur's ends, and so on.

A second problem with the debt-based retributivist solution arises from a case that also caused problems for the fairness-based retributivist solution: the case of the previously victimized offender. McDermott's argument turns crucially on the claim that the offender is currently enjoying a moral good that, if it could be transferred, he would owe to the victim. Because Barry is enjoying a moral good that he would be obligated to transfer to Arthur if he could, for example, it is supposed to be morally permissible for the state now to deprive Barry of the good, even though in doing so it cannot return the good to Arthur. But now consider the following case: Barry is a member of a despised ethnic minority who is constantly having his rights violated by people who tell him that they think of him as subhuman. He is not enjoying the good of being "distinguished as a member of the moral community." Barry then robs Arthur. On the debt-based retributivist account, it is supposed to be permissible now to act in ways that would deprive Barry of some of his rights because this will amount to taking away from him something he would have owed to Arthur: the value of being recognized as a member of the moral community. But Barry does not have this good. And so, on this account, there will be no justification for punishing Barry.

Finally, and perhaps most fundamentally, the debt-based retributivist solution is unacceptable because even if it justifies the permissibility of doing an act that harms the offender, it cannot justify the permissibility of punishing him. Even if McDermott's argument can justify the claim that it is permissible to harm people "as a means of denying these forfeited moral goods to the wrongdoers" (424), that is, it still cannot provide a solution to the problem of punishment. To see that this is so, it may help to return to an example that McDermott uses to develop his argument: the case in which Barry steals something from Arthur and Arthur then dies before Barry can be made to return it (422). In this case, surely, it would be morally permissible to deprive Barry of the use of the material good in question even though this would not help him pay his debt to Arthur. And McDermott claims that in the same way, it would be morally permissible to deprive Barry of a moral good for the same reason. But now consider what would actually be involved in a judge's decision to deprive Barry of the good in question in the first case. Perhaps the judge would order the good to be given to one of Arthur's descendants, or donated to a worthy cause supported by Arthur, or handed over to the state. In all of these cases, the judge would do an act – depriving Barry of the use of the material good – that would harm Barry. But in this case, the judge would not be acting with the intention of harming Barry. He would

simply be taking away from Barry something that Barry was not entitled to have while foreseeing that this would harm Barry. As we saw in Section 1.1.4, punishment must involve not merely harm to the offender but intentional harm. And so, it should be clear that an act that is done to take from someone a good that he does not have the right to use is not a punishment. The debt-based retributivist solution attempts to justify depriving offenders of certain moral goods in precisely this manner. And so, even if it is able to justify this, it will still be unable to justify punishment.

3.4.3 Revenge-Based Retributivism

I have said nothing about the subject of revenge in this chapter. To most retributivists, this is precisely as it should be. The association of retributivism with the feeling of revenge, they complain, is a distortion of the retributivist position, if not an outright smear against it. One defender of the retributivist solution has complained that "[t]he most common way of misunderstanding retributivism is to take it to be a fancy word for revenge" [Gerstein (1974: 76)], and another has dismissed the thirst for revenge as "a primitive emotion which is as vigorously attacked by retributivists as it is by others" [Gendin (1970: 1)]. But for some retributivists, revenge is essential to solving the problem of punishment, and I will therefore conclude my case against the retributivist solution by briefly examining the possibility that there is a satisfactory revenge-based version of retributivism.

Suppose that one wishes to defend punishment by appealing to revenge. How, might one proceed? One possibility is to define revenge as a desire to bring about a state of affairs that is justified on grounds other than revenge. Barton, for example, whose book is presented as a defense of the justice of revenge, seems to proceed in this manner. He treats revenge as the desire that an offender suffer the harsh treatment he deserves, and then he appeals to the legitimacy of revenge in attempting to justify punishment (1999: e.g., 7, 10). But if revenge is construed as a desire to have an offender suffer a deserved form of treatment where the offender deserves to suffer such treatment for reasons other than the fact that the victim (or others) desires that he suffer it, then the desire for revenge itself will play no role in the justification of punishment. If the vengeful desire to see the offender suffer is legitimate because the offender deserves to suffer, that is, then punishing him will be permissible simply because the offender deserves to suffer, and it will not matter whether or not anyone desires to see him suffer.[67] If the desire to see an offender suffer is legitimate because his suffering would be fair, on the

[67] This, in the end, does seem to be Barton's position, and so his book seems in the end to represent a defense of desert-based retributivism.

other hand, then the position that might initially appear to be based on revenge will really be based on fairness, and the same will be true for any other version of this position. If revenge is analyzed as a desire that an offender suffer a treatment that is itself justified on other grounds, in short, then revenge cannot serve as the basis for a retributivist justification of punishment.

The alternative is to define revenge in a way that does not presuppose the permissibility of the state of affairs that the vengeful person seeks, and to argue from that conception to the conclusion that punishment is permissible. Revenge, on this account, is simply the desire to see wrongdoers suffer. Because this conception of revenge does not build the legitimacy of its object of desire into its definition, it does not beg the question by assuming that punishment of the offender is justified. But, for this very reason, it raises a different worry: if revenge is simply a desire that the offender suffer for his offense, how can it be used to justify punishment? How can the merely factual claim that a certain desire exists be used to ground a normative claim about the permissibility of performing acts that would satisfy that desire?

One possible answer to this question is as follows: because the feeling of revenge is so widespread and so powerful, it is virtually inevitable that if the state does not punish an offender, someone else will. The choice, on this account, is not between punishment and no punishment, but rather between punishment by the state and punishment by victims and those who are moved to act on their behalf. Since the state is in a position to judge and mete out punishment impartially, while victims and their friends are more likely to act indiscriminately and disproportionately, it follows that the harms caused by the pursuit of revenge can be minimized by channeling this desire through the state rather than by letting it run wild on its own. This is the view of punishment most widely associated with the famous Victorian judge and legal theorist James Fitzjames Stephen, who, in one of the most widely repeated observations about the nature of punishment, famously declared that "The forms in which deliberate anger and righteous disapprobation are expressed [in the execution of criminal justice] stand to the one set of passions in the same relation which marriage stands to [the sexual passions]" [cited by Murphy (1988a: 3)].[68]

[68] Although Murphy himself seems to construe Stephen as defending the first version of the revenge-based position, on which the desire for revenge is, by its nature, the desire for harms that are justified on grounds other than revenge. As Murphy puts it, "Stephen's point is a simple one: Certain wrongdoers quite properly excite the resentment (anger, hatred) of all right-thinking people, and the criminal law is a civilized and efficient way in which such passions may be directed toward their proper objects, allowing victims to get legitimate revenge consistently with the maintenance of public order" [(1988a: 3–4); see also (1988b: 94–5)].

If a defender of punishment appeals to the desire for revenge in this manner, then she may be able to develop a defense of punishment based on revenge. But this will be of no help to the proponent of the retributivist solution in particular. On this account, after all, revenge will justify punishment only insofar as it can help to demonstrate that the beneficial consequences of punishment outweigh the negative ones. The argument of such writers as Stephen, then, could at most point to a particular instance of the nonutilitarian version of the consequentialist solution discussed in Section 2.4. Since that position was also rejected in Section 2.4 [indeed, the argument of Perkins (1970) discussed there is in effect the same as the one considered here], there is no need to say anything further about it here.

There is, however, a need to say something further about one final way in which a defender of punishment might attempt to develop a viable version of revenge-based retributivism. On this third alternative, a victim's desire for revenge renders the offender's punishment permissible because the offender is responsible for having generated this unsatisfied desire. Because the offender has fulled the victim's desire for revenge, and because the victim is worse off having this desire and having it unsatisfied than she was before the offender created it, the offender now owes it to the victim to satisfy this desire. And since the desire can only be satisfied by having the offender punished, it follows that the offender owes it to the victim to undergo punishment.

This final attempt to develop a revenge-based version of retributivism seems to be the most promising. It also seems to be unsatisfactory. But since the reasons for this involve, at least in part, features of the theory of pure restitution that will be discussed in Chapter 5, I will defer this discussion until the theory itself has been presented.[69] For now, then, I conclude (conditionally, on the assumption that my objection to this final version of revenge-based retributivism proves to be successful) that the retributivist solution to the problem of punishment is, in the end, no more successful than the consequentialist solution. Consequentialism and retributivism are by far the two most prominent responses to the problem of punishment, so this result poses a serious problem for the defender of punishment. But they do not exhaust the responses that have been developed in the contemporary literature on punishment. I will therefore turn, in the next chapter, to various other solutions to the problem, solutions appealing to such considerations as consent, reprobation, moral education, and self-defense, as well as solutions that involve hybrids of other solutions. I will argue that these solutions, too, are unsuccessful and that we have, as a result, no successful solution to the problem of punishment.

[69] See Section 5.13 for the rebuttal of this final version of revenge-based retributivism.

4

Other Solutions

4.0 OVERVIEW

The consequentialist solution to the problem of punishment is forward-looking. It explains the moral permissibility of harming someone now in terms of the positive consequences that it can be presumed to bring about in the future. The retributivist solution to the problem of punishment is backward-looking. It explains the moral permissibility of harming someone now in terms of the wrongness of what the person did in the past. I have argued that both solutions are unsuccessful. If I am correct, then the defender of punishment would seem to have only three remaining options. One option is to identify some other consideration that could justify harming people who break the law. This option will be examined in the first two sections of this chapter. In Section 4.1, I will consider the view that punishment is morally permissible because offenders consent to being punished, and in Section 4.2, I will consider the view that punishment is permissible because society has a right to express its disapproval of unlawful behavior. A second option is to deny that punishment ultimately harms the person punished. This option will be examined in Section 4.3, where I will consider the moral education theory of punishment, on which punishment is justified because it ultimately benefits the person punished. The final option is to concede that punishment ultimately harms the person punished, and that no single forward- or backward-looking consideration can render it permissible, but to maintain that some suitable combination of forward- and backward-looking considerations will prove sufficient. This approach will be discussed in the final two sections. In Section 4.4, I will consider the view that the permissibility of punishment is grounded in the same combination of forward- and backward-looking considerations that permit one individual to harm another in self-defense. And in Section 4.5, I will consider the claim that the permissibility of punishment can be established by

developing a hybrid solution that combines elements of the consequentialist and retributivist solutions in particular or, more generally, that combines elements of any of the solutions considered in this book. I will argue that all of these attempts to solve the problem of punishment are also unsuccessful. The cumulative result of this chapter, along with the two that preceded it, therefore, is that we have no successful solution to the problem of punishment. Virtually everyone believes that the state has the right to punish people for breaking the law, but it turns out that we have no good reason for accepting this belief. This conclusion will, in turn, set the stage for the final chapter, where I will consider what we should do if we have no solution to the problem of punishment.

4.1 THE CONSENT SOLUTION

Larry and Moe are standing on opposite street corners, each hoping to catch a ride home at the end of a long day's work. A car pulls up in front of Larry and stops. Larry gets in and says he would like to go home. The driver takes him home. A second car pulls up in front of Moe and stops. Moe gets in and says he would like to go home. The driver takes him home, too. Larry's and Moe's experiences are identical except for one fact: the car that Moe gets into is driven by his friend, Curly, while the car that Larry gets into is a cab driven by a taxi driver. This difference seems to be morally relevant. Although neither Larry nor Moe has said that he intends to pay the driver for the ride, it seems clear that Larry now owes money to his driver, while Moe does not. Larry and Moe both got into their cars with the right to retain all of their money, that is, but while Moe still has this right after he gets out of Curly's car, Larry does not still have this right as he gets out of the taxi cab. His driver is now entitled to keep some of Larry's money.

The claim that the taxi driver has a right to some of Larry's money, while Curly has no right to any of Moe's money, seems relatively uncontroversial. And the claim itself seems securely grounded in considerations that are neither consequentialist nor retributivist in nature. It is not that the consequences of Larry's paying the driver are positive, while the consequences of Moe's paying Curly are negative. Even if more good would be produced by Moe paying for his ride than by Larry paying for his, we would still believe that Larry, but not Moe, owes money to his driver. And it is not that Larry, but not Moe, owes money because Larry has done something morally worse than Moe has done by getting into a taxi cab rather than into a friend's car. There is nothing wrong with getting into a taxi cab, just as there is nothing wrong with getting into a friend's car. What grounds the morally relevant difference between Larry and Moe in this case is something fundamentally different from either of the kinds of moral consideration that we have examined as possible justifications for the permissibility of punishment. What grounds the difference is that Larry,

but not Moe, has consented to pay for his ride. It is true, of course, that Larry has not explicitly agreed to pay for his ride. But the act of getting into a taxi cab can plainly be considered to involve tacit consent to pay in a way that getting into a friend's car plainly cannot. There are, moreover, many other contexts that seem to have the same feature: raising your hand at an auction, placing your chips on red before the roulette wheel is spun, leaving some money on the table before you leave a restaurant, and so on. This suggests that a solution to the problem of punishment that is neither consequentialist nor retributivist might well arise from a principle of tacit consent. If offenders can be understood as consenting to their punishment (or at least to their liability to being punished), after all, then the fact that punishment involves treating them in ways that would otherwise be morally objectionable need not count as an objection to it.

4.1.1 Punishment and Consent

This consent-based solution to the problem of punishment has been defended most forcefully by C. S. Nino in his article "A Consensual Theory of Punishment" (1983) and later in his book *The Ethics of Human Rights* (1991). I will therefore focus my discussion of this solution on Nino's defense of it. Nino's argument begins with the assumption that consent is, in fact, given in such cases as riding in a taxi or raising one's hand at an auction. It then asks in virtue of what facts this should be so. Nino's answer to this question is twofold. First, consent is given in part because the acts in question are voluntary. If someone forces you into a taxi cab at gunpoint, for example, or raises your arm against your will at an auction, you have not consented to anything. Second, consent is given in part because the voluntary acts have a particular consequence and are done with the knowledge of this consequence. The consequence of the act is normative, not factual. The consequence of getting into a taxi, for example, is not that the driver will take some of your money at the end of the ride. Your getting into the cab is consistent with his later forgetting to ask you for the money or deciding to give you a free ride. Rather, the consequence is that the driver will have the right to take some of your money at the end of the ride. If an act does not have a normative consequence, or if it is done without a suitable understanding of the consequence, then, again, the act cannot count as consenting to the consequence. If, for example, a person who has never heard of taxi cabs accepts a ride from a cab driver and is never told that he is expected to pay for it when the ride is over, then it would be unreasonable to insist that he has consented to pay for the ride.[1] If an act is

[1] This is not to insist that it is obvious that he has no obligation to pay for the ride. Such an obligation might be justified by appealing to some other consideration, such as a principle

done voluntarily, however, and with the knowledge that it has a particular normative consequence, then consent to that normative consequence has been given. As Nino puts it, "the person who voluntarily performs an act knowing that it has the undertaking of certain obligations as a necessary consequence consents to undertake those obligations" (1983: 100).

To this general theory of consent, Nino then adds a further claim about violations of the law in particular: "A necessary legal consequence of committing an offense is the loss of immunity from punishment that the person previously enjoyed" (102). And from this, it seems to follow that a person who voluntarily commits an offense and understands that he is doing so consents to his loss of immunity from punishment. This is not to insist that the person who breaks the law consents to being punished. Whether or not an offender is actually punished is a factual matter, and on Nino's analysis the agent consents to the normative, not the factual, consequence of his voluntary act. Rather, the person who breaks the law consents to the resulting state of affairs in which punishment is now permissible. As Nino concludes his argument, "The individual who performs a voluntary act – an offense – knowing that the loss of his legal immunity from punishment is a necessary consequence of that act consents to that normative consequence" (102). And since the offender has consented to losing any claims against being punished, it is morally permissible for the state to punish him even though this involves treating him in ways that would otherwise be impermissible.

The consent solution to the problem of punishment is undeniably attractive. It does not justify punishing the offender by pointing to the fact that punishing her will create more security for society, or a fairer distribution of goods, or a state of affairs in which people get what they deserve. And so, unlike many of the other solutions that we have examined, it does not objectionably treat people as if they were merely a means of making the world a better place. As Nino emphasizes in defending punishment on consent-based grounds, "we rely on the moral autonomy of the individual, making his liability to punishment depend on his free and conscious undertaking of it" (111). And since virtually everyone will agree that consent is, in fact, given in the sorts of cases that Nino appeals to (the taxi ride, the auction, etc.), the consent solution seems to be well grounded in an uncontroversial moral foundation. The question, then, is whether, given all of this, there is still sufficient reason to reject the solution.

of fairness or utility. The point is simply that such an obligation could not reasonably be justified by appealing to the claim that this passenger has consented to pay merely because he got into the taxi cab voluntarily.

4.1.2 The Moral versus Legal Rights Objection

The answer is that there are several reasons. The first reason arises from a general question that any version of the consent solution must answer: should the claims to immunity that the offender is said to have consented to waive be understood as legal claims or as moral claims? Is the "normative consequence" of the offender's act, that is, a legal or a moral consequence? The question is simple, but either answer to it leads to intractable problems for the consent solution. Suppose, first, that the claims in question are understood as legal claims. This is certainly how Nino himself presents the position. He describes the "loss of immunity from punishment" as a "necessary *legal* consequence of committing an offense" and describes the "normative consequence" that the offender consents to as "the loss of his *legal* immunity from punishment" (1983: 102, emphases added). If this is how the consent solution is understood, then its foundation certainly seems to be reasonable. It is difficult to deny, for example, that murderers, lose some of their legal rights. But, and perhaps precisely because it is so reasonable, this version of the claim about the consequence that the offender consents to will prove unable to justify the conclusion that a solution to the problem of punishment requires. For the problem of punishment is not the problem of how a murderer could come to lose some of his legal rights, but rather how he could come to lose some of his moral rights. The question is not how legal punishment could be legal, that is, but rather how legal punishment could be moral. If a murderer consents to waive his legal right to life but retains his moral right to life, for example, then his consent to waive his legal right to life will support the claim that it is legally permissible to execute him, but it will do nothing to support the claim that it is morally permissible to do so. And the same would be true of all other offenders: establishing that they have lost certain legal protections cannot establish that it would be morally permissible to punish them unless it can also be established that they have lost the corresponding moral protections. This first construal of the normative consequences appealed to by the consent solution is therefore unsatisfactory. It would permit the solution to pass the foundational test but not the entailment test.

For the proponent of the consent solution, however, the alternative construal of the normative consequence of the offender's act leads to the opposite problem: it would permit the solution to pass the entailment test but not the foundational test. Suppose that the consent solution tries instead to build on the claim that the normative consequence of committing an offense is the loss of certain moral claims to immunity. If this is the claim that the consent solution attempts to build on, then it is much easier to see how one could move from the claim to the solution's final conclusion that it is morally permissible for the state to punish people for

breaking the law. If a particular offender has agreed to lose certain moral protections, after all, then it is much easier to see how what she has consented to could render it morally permissible to treat her in ways that would otherwise be immoral. But the problem with this version of the foundation for the consent solution is that appealing to this claim in attempting to justify the moral permissibility of punishment simply begs the question at issue. The question posed by the problem of punishment is precisely the question of why breaking a just and reasonable law should make punishment morally permissible. Therefore, a solution to the problem that assumes that a normative consequence of the offense is a loss of moral immunity to punishment cannot be used. Rather than providing a firm foundation on which to build a defense of the moral permissibility of punishment, this claim simply assumes the moral permissibility of punishment.[2] And, in any event, an appeal to the claim construed in terms of moral rather than legal consequences would render the appeal to consent superfluous. If breaking the law carries the necessary normative consequence that the offender loses certain moral rights, after all, then this will justify the permissibility of punishing her whether she agrees to accept the loss (or even knew it would occur) or not. Whether the rights in question are understood as moral or legal rights, therefore, the consent solution cannot provide a satisfactory basis for solving the problem of punishment. When understood as moral rights, the solution passes the entailment test but not the foundational test; when understood as legal rights, it passes the foundational test but not the entailment test.

4.1.3 The Ignorant Offender Objection

The objection based on the distinction between moral and legal rights poses a general problem for the consent solution. I see no satisfactory way for proponents of the solution to get around it. Even if they do, however, there are a variety of other, more specific problems that will still render the solution unacceptable. The first problem arises from the fact that even if considerations of consent can justify the moral permissibility of punishing some lawbreakers, such considerations cannot justify punishment in all cases. There are two cases in which the consent solution cannot justify punishing the guilty. Since punishment involves the state's right to

[2] Rather than assuming that the loss of certain moral rights is a necessary consequence of breaking the law, one could offer an argument for such a claim. This, in effect, is what the forfeiture-based retributivist attempts to do using the rights-duties argument (see Section 3.2.1). But this strategy seems to amount to abandoning the consent solution in favor of the forfeiture-based retributivist solution.

punish all those who break the law, each case shows that the consent solution fails the entailment test.

The first case that poses this problem for the consent solution arises from the second of Nino's two requirements: that the offender know that his act involves the loss of immunity from punishment. Given this requirement, it follows that the consent solution cannot justify punishing an offender who did not know that his act was illegal.[3] Indeed, Nino recognizes that this result is entailed by his position. On the consent approach to the problem of punishment, he notes, an offender "must have known that the undertaking of a liability to suffer punishment was a necessary consequence of such an act. This obviously implies that one must have knowledge of the law" (1983: 104). But while Nino clearly recognizes this implication of the consent solution, he does not seem to recognize that the implication is a problem for the solution.[4] For while ignorance of facts may sometimes excuse a person from liability to punishment (e.g., a person who accidentally takes another person's coat home from work, genuinely believing it to be her own), ignorance of the law is generally considered an insufficient excuse, and must be so considered by anyone defending punishment for breaking the law. A man who rapes his wife, for example, cannot evade prosecution by saying that he honestly believed that husbands had the legal right to rape their wives, even if he was raised in a culture supporting this belief. A driver cannot avoid being fined for making an illegal right turn at a red light by saying that he believed he had not yet crossed the border separating a state that permits such turns from one that forbids it, and this is so even if it is not his fault that he did not notice that he had crossed the border (perhaps because the small sign announcing this was blocked by a large truck in the other lane as he passed it a few minutes earlier). A person who honestly believed that the law permitted him to kill a clearly unarmed trespasser on grounds of self-defense cannot avoid punishment if the law says otherwise.

[3] In some such cases, this implication may not be unacceptable. If the state prohibits a certain behavior but does nothing to publicize this fact, a defender of punishment might legitimately complain that he is not committed to the state's having the right to punish people who violate the prohibition and that it is therefore no objection to the consent solution that it fails to justify punishment in such cases (indeed, if anything, its failure might well prove a virtue). The assumption that a law is just and reasonable, that is, might be taken to include the assumption that the law has been reasonably well publicized. I will assume in the examples that follow, therefore, that the laws in question have been reasonably well publicized. In these cases, the consent solution will still fail to justify the right to punish an ignorant offender and, in these cases at least, a defender of punishment must acknowledge that this implication is unacceptable.

[4] Few critics of the consent solution seem to have recognized this either, though see Kershnar (2001: 44, 46).

People who break the law are often ignorant of precisely what the law forbids. Indeed, people in general are often ignorant of this, and their ignorance is easily understood. As Reiff, for example, has recently observed, "We may use ordinary language to define many offenses, but give these words highly technical and restrictive legal meanings that are less widely disseminated and are in any event difficult for the general public to understand" (2005: 178). This seemingly inescapable feature of any complex legal system causes a serious problem for the consent solution because ignorance sometimes means not knowing that a certain behavior is illegal.[5] But the problem caused by the ignorant offender here is actually much more far-reaching. For while some people do not know that their acts are illegal, presumably far more people do not know the penalties for the illegal acts they knowingly perform. This poses a further problem because, under the consent solution, in such cases the offenders could be punished only with the amount of punishment that they themselves knew to be attached to their offenses. As Nino himself puts it in characterizing the consent solution, "When that *particular* legal consequence of the voluntary act is known by the agent, we may say that he has consented to it. And it is that consent which is taken to be morally relevant and to justify enforcing the normative consequence in question against the person who has consented to it" (1983: 101, emphasis added). This means that if an offender knew that possessing a certain drug was illegal but mistakenly believed that its possession was a minor misdemeanor rather than a serious felony, then the state would have the right to punish him, but only in the amount attached to a minor misdemeanor. If the law says that possession of the drug renders carries a prison sentence of twenty years, for example, while the offender himself mistakenly believed it mandates only a $500 fine, then under the consent theory, the state only has the right to give him a $500 fine.[6] Since many offenders are

[5] Another example of a largely unrecognized legal prohibition received more publicity when Martha Stewart was convicted of violating "a little-known federal law," often referred to as "Rule 1001" (after the section of the federal code that contains it), that "prohibits lying to any federal agent, even by a person who is not under oath and even by a person who has committed no other crime." Rule 1001 "has become a crucial weapon for prosecutors," but since most people who are charged with violating it are not aware that lying to a federal agent even when not under oath is a federal offense, the consent solution would be unable to justify its use in typical cases [Berenson (2004: Sec. 4, p. 14)].

[6] A defender of the consent solution might maintain that this version of the problem arises only from Nino's formulation of the solution, not from the basic idea underwriting the solution itself. A person who voluntarily gets into a taxi cab, she might point out, does not know what the final fare will be, but he consents to pay it nevertheless. Similarly, she might say, as long as the offender knew that there would be some penalty for possessing an illegal drug, he can be taken to have consented to the penalty, whatever it turns out to be. But this open-ended version of the consent position would produce unacceptable results in both cases: if the individual was unaware that the fare involved forfeiting all of his future

likely to be ignorant of the specific punishment for their particular offenses, it follows that the consent solution cannot justify the appropriate punishment in many cases. And, again, the kind of ignorance that generates the problem here is not uncommon. A 1998 study in Great Britain, for example, found that a large percentage of those surveyed significantly underestimated the the typical penalty for a variety of offenses [Reiff (2005: 184fn10)].

A defender of the consent solution might try to respond to the ignorant offender objection in two ways. One would be to deny that the solution has this implication. Even though the husband who rapes his wife might have genuinely believed that he had the legal right to do so, for example, a defender of the consent solution might argue that he should have known better. He was ignorant of the law, that is, but he was nonetheless culpable for his ignorance. Culpable ignorance is a potentially perplexing problem, but its various nuances need not concern us here for two reasons. First, there are too many laws to insist plausibly that one is responsible for knowing all of them and, as already noted, the laws often use terms in a highly technical sense that is difficult for nonspecialists to fully understand. So, it is implausible to suppose that every time a person does an act without realizing it is illegal, it is her own fault that she did not know this. Second, and more fundamentally, even if we agree that a particular offender is culpable for her ignorance of a particular law, this cannot help to salvage the consent solution. Knowledge of the law is still necessary for a voluntary act to count as an act of consent on that solution; so, all that will follow from this will be that, for example, it is the husband's fault that raping his wife did not count as consent to be punished. And surely that is not enough. The same, moreover, will be true of cases involving ignorance of the penalty for a particular offense. It would be sensible advice, for example, to say that if you are going to purchase illegal drugs, then you should find out how they are classified by the legal system. A person who knows that the drugs he buys are illegal but doesn't bother finding out how severe the penalties are is foolish. But if he mistakenly thinks that he could only get a modest fine if he is caught with a particular drug when in fact the penalty is up to twenty years in prison, then even if it is his own fault that he is ignorant of the law, his voluntary act of buying the drug still cannot be understood as his consenting to render himself liable to twenty years in prison.

The other option for the defender of the consent solution at this point is to bite the bullet and admit that people should be punished only if they knew that their acts were illegal and should receive only as much

earnings, for example, or that the penalty for drug possession was being slowly tortured to death, then we would have to say that the individuals in question had consented to these consequences, too, and this is surely implausible.

punishment as they were aware they could receive. But this option is also unacceptable. In the first place, the result is too strongly counterintuitive. Most defenders of punishment want to justify not just punishment but punishment that is intuitively proportionate to the offense. And making offenders subject only to the level of punishment that they believed they would get would render this impossible. This means that, for most people, the problem posed by the ignorant offender shows that the consent solution fails the foundational test. In addition, though, and more fundamentally, this response amounts to abandoning the claim that punishment is justified in the first place, since punishment involves punishing people for breaking the law and ignorant lawbreakers are people who break the law. The ignorant offender objection, therefore, successfully demonstrates that even if consent justifies punishment in some cases, it doesn't do so in enough cases. And if it doesn't do so in enough cases, then it cannot pass the entailment test even if it does pass the foundational test.

4.1.4 The Explicit Denial Objection

There is a second case in which the consent solution fails to justify punishment of the guilty, and this case is, if anything, even more damaging than the first. The problem arises because, when we believe that voluntary actions can amount to tacitly consenting to something, we always believe that the presumption of consent can be overridden by an explicit declaration to the contrary. Suppose, for example, that a doctor who is on call is standing by the roulette table in a casino when his pager goes off. In order to get his pager out, he must first remove several hundred dollars worth of chips from his pocket, and in order to look at the pager to see the message, he must put the chips down. Since the message may be urgent, he puts the chips down on the nearest surface, the space on the table where one would place one's chips in order to bet that the ball will land on red. But since he intends only to put the chips down so that he can check his message, not to place a bet on red, he clearly announces his intention to everyone present, making it clear that he is not placing a bet on red. In this case, if the roulette wheel is spun and the ball lands on black, it would be absurd to say that the doctor had consented to lose his money. The result is that a third requirement must be added to the two that Nino has provided in his analysis of the necessary conditions for tacit consent: for an act to count as tacit consent, it must not be accompanied or preceded by an explicit declaration that one is not consenting. This requirement is needed to avoid unacceptable results in cases such as that of the doctor on call, and one could construct similar stories about people getting into taxi cabs, lifting their arms at auctions and so on. If, for example, you ask a taxi driver to take you

home, but inform him before he begins to drive that you are simply asking him for a favor, that you will not pay him, and that your accepting a ride from him cannot be taken as an agreement to pay him, then it would be plainly implausible to insist that if the driver takes you home, you have consented to pay him. But once this further requirement is added to the analysis of tacit consent, this further problem with the consent-based solution becomes clear: any offender could free himself from liability to punishment merely by announcing that in doing a certain act he did not intend to consent to liability to punishment.[7] This result will surely strike most defenders of punishment as intuitively unacceptable, and this suffices to show that the consent solution fails a reasonable application of the foundational test. And independent of its intuitive plausibility, this result prevents the consent solution from drawing the line between those who may be punished and those who may not in the way that punishment requires, and so prevents the consent solution from passing the entailment test.

4.1.5 The Punishing the Innocent Objection

I have argued that even if the consent solution can justify the permissibility of punishment in some cases, it fails to do so in other cases and fails to justify the permissibility of enough punishment in still others. But let us now suppose that I have been mistaken, and that the consent solution would justify the right to punish everyone who breaks the law and to do so in just the right amount. Even if this is so, there is another problem. For not only will the consent solution justify punishment in all cases in which a successful solution should justify it, but it will also justify punishment in many additional cases in which it should not. Specifically, the consent solution will render punishment permissible in some cases in which a person has broken no laws at all and in many cases in which a person has broken an unjust law. I will consider the objections that arise from these two cases in turn.

The punishing the innocent objection arises from a further shortcoming in Nino's analysis of the conditions for tacit consent. On Nino's account, tacit consent requires the agent to know that his act has a certain normative consequence. For a person to *know* that his act has a certain consequence, however, his act must, in fact, have this consequence. And this renders Nino's requirement too strong. Whether or not a person has

[7] A defender of Nino's account here might maintain that since the normative consequence of losing one's immunity to punishment is said to be a "necessary" consequence of breaking the law, it follows that one cannot prevent the consequence from occurring merely by attempting to deny it. If the consequence follows from the agent's act regardless of the agent's intention, however, then it also cannot matter whether or not the agent is aware of this fact, and the consent solution will therefore collapse into the forfeiture-based version of retributivism considered and rejected in Section 3.2.

consented to something is a function of the person's state of mind; so, while it matters whether or not he believes that his act has a certain normative consequence, it does not matter whether or not his belief is correct.

Suppose, for example, that Larry is eating in a restaurant in a foreign country. At the end of his meal, he puts some money on the table as a tip for the waiter. He believes that in this country, leaving money on the table is understood to amount to consent to transfer the right to the money to the waiter. And, if we like, we can stipulate that he came by this belief in a perfectly reasonable manner. Several knowledgeable people had recommended a particular guidebook as the most reliable source of information about this country, and the book's sole error turned out to be the claim that tipping is a familiar convention there. Larry leaves the money on the table because he believes that it will be understood as his tacit consent to transfer the right to the money to the waiter. Nonetheless, it turns out that Larry is mistaken. As a result, after he leaves the restaurant, the money is treated as lost or abandoned goods.

There are two ways in which we could characterize Larry's act. We could say that he did indeed consent to give the waiter his money, but that the waiter failed to understand this, or we could say that Larry attempted to consent to give the waiter the money but failed to do so. But the latter description is surely strained and unmotivated, and there are three good reasons to prefer the former. First, to consent to something is to agree to it, and whether or not a person has agreed to something cannot be a function of whether or not other people recognize this. As far as Larry is concerned, he has clearly agreed to let the waiter keep the money even if the waiter does not recognize this fact. Second, consider how Larry would respond if he learned later that the waiter did, in fact, take the money. Clearly, he would not believe that he had any basis for complaint about this result. That, after all, is why he put the money on the table in the first place. But if Larry did not actually consent to let the waiter keep the money, then it is just as clear that he would have a basis for complaint: if someone takes your money without your consent, that is a basis for an objection. Finally, consider what we would say in a parallel case involving explicit rather than tacit consent. Suppose, for example, that Larry had clearly said, "I hereby consent to transfer my right to this money to the waiter," but that the waiter did not speak English well enough to understand what Larry meant. Clearly, in this case we would say that Larry had, in fact, consented to give the waiter the money. We would not be tempted to say that because the waiter did not understand Larry's words, Larry had not really consented in the first place. But what goes for explicit consent must surely go for tacit consent, and so, for all of these reasons, we must conclude that Nino's account of tacit consent requires revision. What should be required for consent is not that the agent know

that his act has a certain normative consequence, but that he believe it does (or, perhaps, that he reasonably believe it does).

This further revision is needed to accommodate our judgments about cases such as that of Larry and the foreign restaurant. However, once the revision is accepted, the attempt to use tacit consent to solve the problem of punishment gives rise to the punishing the innocent objection. Suppose, for example, that Moe believes that people who smoke in the restaurant in which he is eating are subject to a $100 fine. If we like, we can stipulate that Moe's belief is reasonably well grounded. He was told this by a person with generally reliable knowledge of the law, for example, or knows from past experience that it is true of other seemingly similar restaurants, and so on. Nonetheless, it turns out either that there is no such law, or that the law was recently repealed, or that this particular restaurant is exempt from the law because of its size, location, or design. Although it is not illegal to smoke in this particular restaurant, Moe (reasonably) believes that it is, and he decides to smoke even though he has this belief. This case is structurally symmetrical to the case of Larry and the foreign restaurant. In both cases, the agent does an act in the (reasonable) belief that it has a certain normative consequence. In the case of Larry, I argued that this was sufficient to establish that he had, in fact, consented to this normative consequence. But if that is so, then we must say the same of Moe.[8] Although he has broken no law, he has consented to be liable to a $100 fine. It therefore follows that under the consent solution, the state could punish Moe by giving him a $100 fine even though he is legally innocent. And although Moe's case is somewhat narrow, it seems plausible to suppose that there are many instances in which people behave in ways that they mistakenly believe to be illegal. Laws concerning such diverse practices as sodomy, statutory rape, drinking, drug use, or simply making a right turn at a red light vary considerably from state to state and somewhat from time to time. There will therefore be many instances in which, unacceptably, under the consent solution, it would be morally permissible to punish someone who has not broken any law merely because he mistakenly believes that he has done so. Because this result will strike most people as strongly

[8] There is, of course, a potentially important difference between the two cases. Larry does his act *because* he believes that it has a certain normative consequence, while Moe does his act *despite* the fact that he believes it has a certain normative consequence. This difference might be appealed to as a way of showing that Larry has consented to the normative consequence of his act, while Moe has not. While this distinction would, indeed, prevent the consent solution from entailing that it is permissible to punish the innocent, however, it would also prevent the consent solution from entailing that it is permissible to punish the guilty since they, too (with some possible exceptions), act despite rather than because of their belief that their act is illegal.

counterintuitive, it will be justifiable to conclude that the consent solution fails the foundational test. And since, in any event, it shows that the solution cannot draw the line between those who may and may not be punished in the place where punishment draws it, the objection clearly demonstrates that the consent solution cannot pass the entailment test.

4.1.6 The Unjust Laws Objection

The punishing the innocent objection is one case in which the consent solution justifies punishing someone who should not be punished. A second case concerns punishing someone who has broken an unjust law. This unjust laws objection can be developed in two ways: appealing to cases in which a law forbids an act that it would be morally wrong to forbid and appealing to cases in which a law provides a morally unacceptable penalty for doing an act that it is morally permissible to forbid. In both cases, the implications of the consent solution are consistent with punishment, so neither threatens the ability of the solution to pass the entailment test. But in both cases, the implications are so strongly counterintuitive that for virtually anyone interested in defending punishment, they show that the consent solution fails the foundational test.

Suppose, to begin with an example of the first case, that you live in a country ruled by a dictator. He has imposed a law under which people are sentenced to twenty years of hard labor in a brutal prison camp for publicly criticizing the government. You are aware of this law. Nonetheless, you publicly and voluntarily criticize the government. Indeed, you publicly and voluntarily criticize the government precisely because it has imposed this law. You are arrested. Nino's position implies that you have consented to waive your liability to punishment, and that it would therefore be morally permissible for the state to condemn you to twenty years of hard labor for having publicly criticized it. And suppose, to borrow an example of the second sort from Alexander, the state unobjectionably forbids people from overparking but objectionably imposes capital punishment on those who overpark. In this case, as Alexander points out, the consent solution will entail that "one who voluntarily overparks 'consents' to be executed" (1986: 179). But this implication is unacceptable. And so, once again, is the consent solution to the problem of punishment.[9]

[9] The unjust laws objection has been endorsed, in one form or another, by a number of writers, including Goldman, who puts the point well: "In general, having warned someone that he would be treated unjustly is no justification for then doing so, even if, once warned, he could have avoided the unjust treatment by acting in some way other than the way he acted" (1979: 43). See also Burgh (1982: 199), Braithwaite and Pettit (1990: 169), and Knowles (1999: 41).

So long as the proponent of the consent solution insists that consent to being liable to punishment gives the state the moral right to punish, there can be no adequate response to the unjust laws objection. The consent solution, given this restriction, clearly has this implication, and the implication is clearly unacceptable. The only possible reply, therefore, would be to deny that consent alone is sufficient to render punishment permissible. This, in the end, is the position that Nino himself finally retreats to. Although at times he speaks as if consent by itself is sufficient to permit punishment [he claims, for example, that it is "consent which is taken to be morally relevant and to justify enforcing the normative consequence in question against the person who has consented to it" (1983: 101)], he is ultimately forced by the unjust laws objection to claim only that consent is necessary but not sufficient for permissible punishment.

> *If* the punishment is attached to a justifiable obligation, if the authorities involved are legitimate, if the punishment deprives the individual of goods he can alienate, and if it is a necessary and effective means of protecting the community against greater harms, *then* the fact that the individual has freely consented to make himself liable to that punishment (by performing a voluntary act with the knowledge that the relinquishment of his immunity is a necessary consequence of it) provides a prima facie moral justification for exercising the correlative legal power of punishing him. [(1983: 104, emphasis added; see also Nino (1986: 184)]

The claim, that is, is no longer that consent is something that can "justify enforcing the normative consequences in question," but rather that the appeal to consent can be used to fill "the gap in the moral justification of the practice, left by pure considerations of social protection" (1983: 104). The agent's consent, that is, is in the end necessary but not sufficient for the moral permissibility of punishment.

But this response to the unjust laws objection is unacceptable for two reasons. First, it seems to be entirely ad hoc. If consenting to be liable to punishment for violating a useful, effective law by breaking it suffices to permit punishment, why shouldn't consenting to be liable to punishment for violating a counterproductive, ineffective law by breaking it similarly suffice? If consenting to pay $10 million for a valuable painting by raising your hand at an auction suffices to obligate you to pay $10 million for it, after all, then surely consenting to pay $10 million for a worthless painting by raising your hand at an auction also suffices to obligate you to pay $10 million for it. If putting your chips on the table means that you have consented to pay if you lose when making a wise choice with your cards, it must equally mean that you have consented to pay if you lose when making a foolish choice.

Second, this response to the unjust laws objection misses the full force of the objection itself. Part of the force of the objection arises from its claim that the consent theory implies that it would be permissible to imprison the political agitator and execute the overparking driver. The response does attempt to overcome this part of the objection by saying that even though these people have consented to be liable to punishment, other considerations would make it wrong to treat them in these ways. But most of the force of the objection arises simply from the fact that it is objectionable to say that these people have consented to be treated in these ways in the first place. The dissident, for example, clearly believes that the law he is violating is unjust. It is absurd to maintain that by voluntarily criticizing the government's legal prohibition he is tacitly acknowledging the legitimacy of the government's right to punish him for violating it. Yet, the consent solution to the problem of punishment, even in its modified form, still has this implication. And since this implication alone seems clearly to be unacceptable, the consent solution again fails to provide a reasonable foundation for a successful solution to the problem of punishment.

4.1.7 The Harm versus Punishment Objection

I have argued that the consent solution, in effect, justifies both too little and too much: too little in that it fails to justify the right to punish people when a defender of punishment will say that they may be punished, and too much in that it succeeds in justifying the right to punish people when a defender of punishment will say that they may not be punished. Even if I have been mistaken about all of this, however, there is still one final problem with the consent solution. Even if it justifies harming those and only those who violate just and reasonable laws, it does not justify intentionally harming them for breaking those laws. Since punishment involves not merely harm but intentional harm, it follows that the consent solution fails to justify punishment even if it succeeds in justifying harm. And so, for one final reason, the consent solution must be judged to fail the entailment test that any satisfactory solution to the problem of punishment must pass.

The problem here can perhaps be seen most clearly by beginning with a simple case in which it is uncontroversial that it is permissible to take something of value from someone because he has tacitly consented to have it taken away. So, suppose that Larry walks into a restaurant and orders a grilled cheese sandwich. He never says explicitly that he will pay for the food, and the waitress who takes his order never explicitly asks him if he will pay for it. Still, given the context, it seems clear that Larry may reasonably be taken to have consented to pay the amount listed on the menu. In this case, it seems clear that even if Larry does not want to

pay for the sandwich after he eats it, the restaurant has the right to collect payment. It seems equally clear that this does not mean that the restaurant has the right to punish Larry. The restaurant does not seek to make Larry suffer for having eaten one of its sandwiches. It simply wishes to take from Larry what he has agreed to have taken from him, whether or not this makes him worse off than he would otherwise be. But now consider that, according to the consent solution, the relation between the restaurant and its customer is precisely the same as that between the state and the offender. Suppose, that is, that a man who has robbed a liquor store has consented to be liable to imprisonment for five years. In that case, on this account, the fact of consent really does give the state the right to imprison him. Locking him up for five years will harm him. And so, on this account, the state has the right to do an act that will harm him. But as with the case of the customer in the restaurant, it does not follow that the state has the right to intentionally harm him. Punishment, as we have noted on a number of occasions, involves not merely harm but intentional harm. And so, even if the consent solution can justify harming offenders, it cannot justify punishing them. Therefore, even if the consent solution can overcome all of the objections that I raised against it in the previous sections, it will still fail the entailment test and will not, for one final reason, be a satisfactory solution to the problem of punishment.

4.2 THE REPROBATIVE SOLUTION

A given act is an act of punishment, in part, because it is intentionally harmful. This is why punishment raises a moral problem. But there is more to punishing a person than harming her. In particular, as we saw in Section 1.1.6, there is a reprobative element to punishment. Punishing an offender, that is, involves conveying an attitude of censure, disapproval, condemnation. Most attempts to solve the problem of punishment set this feature of punishment aside. But there is one solution to the problem that makes this feature of punishment its focus. This is the reprobative solution, on which the state has the right to punish offenders precisely because it has the right to convey its disapproval of the offender. As Hart puts the point, "What distinguishes a criminal from a civil sanction and all that distinguishes it ... is the judgment of community condemnation which accompanies and *justifies* its imposition (1958: 14, emphasis added).[10] The claim that the right to punish may derive from the right to

[10] The reprobative solution is also nicely captured by Lord Denning's widely quoted defense of capital punishment: "The ultimate justification of any punishment is not that it is a deterrent, but that it is the emphatic denunciation by the community of a crime: and from this point of view there are some murders which, in the present state of public opinion, demand the most emphatic denunciation of all, namely the death penalty"

express disapproval is distinct from all of the defenses of punishment considered to this point. It therefore merits discussion here.

4.2.1 Punishment as Censure

The most prominent and thoughtful contemporary defender of the reprobative solution is R. A. Duff. "Given an appropriate set of conventions and a shared understanding of those conventions," on Duff's account, "a term of imprisonment or compulsory community service, or a fine, can communicate to those on whom it is imposed (and to others) an authoritative censure or condemnation of the crime for which it is imposed. Such punishments are then no longer merely hard treatment, but also symbolic acts of censure" (2001: 29, emphasis deleted). The claim that it is *possible* for the state to censure an offender by punishing her, of course, is not the same as the claim that it is *permissible* for the state do to so. But Duff attempts to bridge the gap from the former to the latter by means of two important further claims: the claim that the state should, in fact, censure those who break the law and the claim that "penal hard treatment is not just a possible, but a necessary, method of communicating the censure that offenders deserve" (29). If both of these claims can be sustained, then the moral permissibility of punishing people for breaking the law can be derived from the moral permissibility of censuring people for breaking the law. I will argue in what follows, however, that the first claim can only be sustained in a sense that is too weak to do the work needed of it and that the second claim cannot be sustained at all.

4.2.2 The Weak versus Strong Rights Objection

The central argument for the first claim, the claim that society should (or at least may) express its disapproval of people who break the law, maintains that the claim follows logically from the fact that what offenders do is prohibited. If a society truly believes in a given legal prohibition, that is, then this commits it to expressing its disapproval of those who

[quoted by Hart (1962: 170)]. See also Miller (1966, 1970), Feinberg (1970), Primoratz (1989b: 187ff.), von Hirsch (1993), Hampton (1988c: 124–38), although Hampton defends this view as a way to defend retributivism, and Metz (2000). In (1988c), Hampton is often concerned to reconcile this position with the New Testament, but as she emphasizes in 1988a (12–13), the argument is not meant to depend necessarily on Christian assumptions. At times it rests on, but can also be made independent of, her account in (1988b). And, most importantly, Duff (1986, 2001). Thoughtful critiques of Duff (1986) can be found in Bickenbach (1988) and Baker (1992); Duff's reply to the former appears in Duff (1988); and an interesting defense and development of Duff's position, focusing on justifying the offender's acceptance of his punishment, can be found in Tudor (2001).

violate it. As Duff puts it, "An honest response to another's culpable wrongdoing ... is to criticise or censure that conduct. ... So, too, a society which declares certain kinds of conduct to be wrong, as criminal, can and should then censure those who nonetheless engage in such conduct" (1999: 50; see also 2001: 28).[11] Indeed, not only is it honest for society to express its disapproval of offenders on this account, but it would be positively dishonest to refrain from doing so: "to remain silent in the face of crime would be to betray the values which the law expresses, and to which we are committed" (Duff 1986: 236).[12] It is, Duff maintains, a "conceptual point" that "[t]he criminal law declares certain kinds of conduct to be wrong – to be criminal. But if the law, or the society in whose name it speaks, is to mean what it thus says, it is committed to censuring those who nonetheless engage in such conduct" (2001: 28).

Now the claim that the state has a particular right can be understood in one of two ways. On a relatively weak reading, the claim means merely that the state has the right as a prima facie right: the right permits the state to do something provided that, in doing so, it does not employ any means that are themselves independently objectionable. It is generally taken to be uncontroversial, for example, that the state has the right to print its own currency, but this does not mean that it can do any act at all so long as it does the act as a means of printing currency. It means merely that so long as nothing involved in its printing of currency is independently morally objectionable, it may print its own currency. The fact that its acts amount to acts of printing currency cannot itself be an objection. On a much stronger reading, the claim that a certain right exists means that behavior that would otherwise be impermissible is rendered permissible by the existence of the right. This is how the right to self-defense is standardly viewed, for example. Normally, it would be impermissible for me to kill you, but if your attacking me activates my right to

[11] As Primoratz puts the argument, "Rules that state standards of behaviour and command categorically imply that actions violating them are wrong, and that such actions are to be condemned, denounced, repudiated. Expressions of this condemnation and repudiation are the index of the validity of the rules and of the acceptance of the conviction that their breaches are wrong in society" [(1989b: 196); see also (1989a: 151)]. The same argument is also endorsed by a number of other writers, including Oldenquist (1988: 467), Tunick (1992: 108), and von Hirsh (1993: 9).

[12] Although the argument for which Duff offers this claim comes closer to the moral education position considered in Section 4.3. In general, Duff can be understood as developing arguments in defense of both the reprobative solution (when he argues in terms of censure) and the moral education solution (when he argues in terms of the cleansing role of penance), but when I refer to Duff's position in this section, I mean to refer to his defense of the reprobative solution. See also Hampton (1988c: 131), who argues that if we don't respond to the offense, "we would be acquiescing in the message it sent about the victim's inferiority."

self-defense, then this right makes it permissible for me to do what would otherwise be impermissible.

This distinction between two senses in which it might be claimed that a certain right exists undermines the reprobative solution to the problem of punishment for the combination of two reasons. The first is that the solution can succeed only if it can justify the claim that the state has a right to censure people who break the law in the strong sense of having such a right. Since legal punishment involves the state's treating people in ways that would ordinarily be impermissible, that is, the claim that the state has the right to censure people who break the law can only show that it has the right to punish them for breaking the law if the right to censure is strong enough to render permissible what would otherwise be impermissible. But – and this is the second reason – the claim that the state has a right to censure those who break the law is plausible only when it is construed in the weak sense. The reprobative solution, therefore, is undermined by an equivocation over what it is for the state to have a right to censure those who break the law.

The argument for the state's right to censure is ultimately an argument from integrity. To return to Duff's formulation, if the law is to "mean what it . . . says," then it necessarily follows that it must be "committed to censuring" those who break the law. But while this argument might plausibly ground a right to censure in the weak sense, it cannot ground a right to censure in the strong sense. And the reason it cannot is that it neglects the distinction between believing a proposition and expressing one's belief in a proposition. That a given proposition is true is always a good reason to believe it. And so, if it is true that what a given offender has done is contemptible, then this is always a good reason to believe that what the offender has done is contemptible. But the mere truth of a proposition is not in itself a good reason to believe that one ought to express one's belief in its truth, nor is it a good reason to believe that it would be morally permissible to express one's belief in it. This is so because expressing one's belief in a proposition is an act, since such acts can have negative consequences and can violate rights even in cases where merely holding the belief does not, and since whether or not an act has negative consequences or violates rights is relevant to what one ought to do and to what one may permissibly do. It follows from this that it takes more to justify the permissibility of expressing a belief than the mere fact that the belief is true. The proposition that a triangle has three sides is true, for example, and its truth is a good reason to believe that it is true. But if expressing the belief on a particular occasion would predictably cause a surgeon to lose his concentration and fail to save the life of the patient she was operating on, then the fact that the proposition is true would not be a good enough reason to believe that one should, or permissibly could, express it. If the only way to express the belief would

require one to steal someone else's paper and pencil, then, again, the truth of the proposition would not suffice to establish that one should, or permissibly could, express it. Nor, in either case, could one credibly say that failing to express the belief that a triangle has three sides would represent a failure of integrity. Your belief that a triangle has three sides, that is, gives you only a prima facie reason to think that you should publicly express that belief, a reason that is easily overridden in these cases.

This distinction between having a sufficiently good reason to believe a proposition and having a sufficiently good reason to express a belief in it is crucial in the context of punishment, in particular, for two reasons. The first is that there are many cases in which expressing the (correct) belief that what an offender did was wrong would cause more harm than good. Consider, for example, a parent whose unlawful negligence in failing to put his young child in a car seat led to the child's death. In this case, it is true that the parent behaved wrongly, and we are therefore entitled to believe this. But assuming that the parent already knows that he has done something wrong and feels terrible about it, it does not follow from the fact that we should believe he has behaved wrongly that we should express this belief publicly. Doing so would seem to be an unwarranted act of rubbing it in, and refraining from doing so would in no way represent a lack of integrity on our part. One need not take an unduly charitable view of those who break the law, moreover, to see that such cases are not uncommon. It seems plausible, for example, to suppose that most drunk drivers who kill someone feel terrible about what they have done. As long as they fully understand that what they did was wrong, therefore, the fact that what they did was wrong will not be enough to justify our publicly expressing this fact.[13]

And, indeed, there can be cases in which an offender already feels *worse* about himself than we think he should feel. Suppose, for example, that Larry steals a dollar from Moe as Moe is walking down the street; that as a result of this, Moe turns around to get some more money from an ATM machine; and when he does so, he is accidentally hit and killed by a bus. In this case, Larry knows that he should not have stolen the dollar from Moe and knows that if he had not done so, Moe would not have been killed by the bus. And so, Larry feels terrible about what he has done. In this case, not only would it seem cruel of us to track down Larry and publicly declare how much we disapprove of him for having stolen a

[13] A defender of the reprobative solution might argue that even if the drunk driver is already repentant, there is still good reason for society to censure him. Censuring him, it might be argued, would reaffirm the value of the victim and would show that society treats the offender as a responsible agent. But while these might be good reasons for society to do something in response to the offender's wrongful act (having him speak to groups about the dangers of drinking and driving, for example), it is not clear why they would count as reasons for censuring him in particular, on the assumption that it is already clearly understood that what he did was wrong.

dollar from Moe, but it might seem that we should instead attempt to comfort him and encourage him not to be so hard on himself. And so, in cases in which the offender already feels as bad as we think he should feel, or perhaps even worse, the fact that he has done something wrong provides no reason to believe that we should declare that his act is wrong. All of this, of course, is perfectly consistent with maintaining that the state has the right to censure people for breaking the law in the weak prima facie sense. For on that account, we can simply maintain that the negative consequences of expressing disapproval suffice to override the prima facie right to express it. But none of this can be rendered consistent with the claim that the state has the right to censure in the strong sense in which the right renders permissible what would otherwise be impermissible. And since the reprobative solution depends on the claim that the state has the right to censure in this strong sense, these considerations suffice to undermine the reprobative solution.[14] The only sense in which the state can plausibly be said to have the right to censure offenders is not strong enough to justify the claim that the state has the right to punish offenders. The reprobative solution therefore fails the entailment test that any satisfactory solution to the problem of punishment must pass.

4.2.3 The Nonpunitive Censure Objection

I have argued that the state has the right to censure people for breaking the law only in the weak prima facie sense, and that the reprobative solution can succeed only if the state has this right in the stronger sense in which a right can render permissible what would otherwise be impermissible. But let us next suppose that I have been mistaken about this and that the state's right to censure is as strong in this respect as a person's right to self-defense. Even if this is true, the reprobative solution must still be rejected. In the case of self-defense, after all, the right makes it permissible to kill someone whom it would otherwise be impermissible to kill, but only if there are no other means of self-defense. If you can safely escape an attacker without harming him, for example, then the right to self-defense does not give you the right to kill him. And this suggests that even if the state has the right to censure an offender in the strong sense, in which the right can render it permissible to treat the offender in ways that would otherwise be impermissible, this will justify the permissibility of punishment only if there are no nonpunitive forms of censure. And it

[14] Relatedly, the reprobative solution will also fail in cases in which it is implausible to believe that the offender's act was wrong. If a person drives his dying friend to the hospital in the only car that is available to him, and that car has not yet passed a required emissions test, then if punishment is permissible, the state has the right to punish him for driving the car. But it is extremely implausible to say that this is because the state has a right to express its disapproval of his act.

seems clear that the state could censure an offender without punishing him. It could, for example, issue an official statement of denunciation, in much the same way that legislative bodies sometimes issue a statement censuring one of their members. At the conclusion of a trial when the offender was found guilty, the judge could make a public proclamation that society disapproves of what the offender has done, the statement could be reprinted in the newspaper, and so on. Even if we agree that the state has the right to express disapproval of the offender in the strong sense of having such a right, therefore, this concession cannot be used to justify the conclusion that the state has the right to punish him, so the reprobative solution will again fail the entailment test.[15]

There are three ways in which a defender of the reprobative solution might respond to the problem raised by the prospect of nonpunitive forms of censure. The first is to deny that anything short of punishment can express society's disapproval of the offense. Primoratz, for example, argues that punishment is necessary for an offender's act to be repudiated, that if the offender is not punished, the wrongness of his act has not been affirmed:

> This condemnation is expressed by punishment. By giving expression to it, punishment vindicates the law broken, reaffirms the right violated, and demonstrates that its violation was indeed a crime. Thus, if there are to be rights sanctioned by the criminal law, if some actions are to be crimes, if there is to be criminal law at all, there *must* be punishment as well. Where there is no punishment, there are no crimes, no criminal law, no rights determined and sanctioned by such law. (1989b: 197, emphasis added)[16]

Similarly, Hampton has argued that punishment "is a symbol that is *conceptually required* to reaffirm a victim's equal worth in the face of a challenge to it" (1988c: 125–6, emphasis added). To express our belief that the offender has wronged his victim, that is, Hampton argues that we must symbolically "defeat" the offender and that "the infliction of pain constitute[s] such a symbol" because "pain conveys defeat" (1988c: 126).

The problem with this first response to the nonpunitive censure objection is the same as the problem with the logical entailment argument examined in Section 1.2.3. This response, like that argument, presumes a false dichotomy: either the state does nothing at all to someone who breaks the law or it punishes him in particular. And while it seems plausible to say that if the state does nothing at all in response to an offense this will leave the offender symbolically "undefeated," it does not

[15] The gap in the argument here has been noted by, e.g., Scheffler (2003: 76).

[16] Similarly, "if thieves *as a rule* were not prosecuted and punished, the conclusion would have to be drawn that theft is not really a crime, and that property rights do not really obtain, at least in the sense of rights determined and guaranteed by the criminal law" (1989b: 197).

follow that this will be so as long as the state leaves the offender unpunished. Suppose, for example, that Larry steals Moe's car and that, in response, the state compels Larry to make restitution to Moe: to return the car to him, to compensate him for the various costs he incurred while the car was missing, to apologize to Moe, and so on. If the state did all of this, no one would have difficulty understanding that the state was saying that Larry had done something he had no right to do. And so, the attempt to respond to the nonpunitive censure objection by insisting that it is logically impossible to censure without punishing must be rejected.

A second and more moderate response to the nonpunitive censure objection concedes that it is possible to denounce an offender without punishing him but maintains that there are reasons to prefer punitive denunciation to nonpunitive alternatives. Von Hirsch, for example, who expresses doubts about the first response, argues that public denunciation can be accomplished by either verbal sanction or physical punishment; that the state's right to denounce the offender therefore gives it the right to do either of these things; and that, since the state has the right to do either of these things, it is permissible to appeal to the deterrent value of punishment as a tie breaker between the two. The positive consequences of punishment play a role in explaining why punishment is preferable to mere verbal censure, on this account, but the "blaming function" of punishment still has "primacy" in the structure of its justification (1993: 12).

The problem with this tie breaker response to the nonpunitive censure objection is that tie breakers count only in cases where all else is equal. Suppose, for example, that the state had the right to paint a big sign saying that the offender was a bad person, and that it had to choose between painting the sign red and painting it blue. Given that there is no morally relevant difference between denouncing someone with a red sign versus a blue one, it would be permissible for the state to choose between the two by, for example, selecting the less expensive paint. And so, if there were no morally relevant difference between verbally censuring an offender (or compelling her to make restitution) and punishing her, then it would be permissible to invoke the presumably positive social consequences of punishing her as a tie breaker to decide between the two. But the problem of punishment arises in the first place precisely because there is such a clear and importantly relevant difference between nonpunitive and punitive forms of censure: the latter involve intentionally harming the offender, while the former do not. And so, the appeal to the deterrent value of punishment as a tie breaker between punitive and nonpunitive alternatives is misplaced. So long as there are nonpunitive means of conveying society's disapproval of an offender's behavior, the state's right to denounce the offender cannot generate a right to punish her.

I have argued that there are nonpunitive ways to denounce an offender and that, if this is so, then the right to censure an offender cannot justify a

right to punish her. But there is one final way that a defender of the reprobative solution might respond. He might insist that to denounce an offender is to punish her. If this is so, then as soon as we concede the right to censure, we already concede the right to punish. This final response arises from the recognition that it is bad for a person to be publicly denounced. If a man is convicted of soliciting a prostitute, for example, and if the state then publishes his name in the newspaper, along with an official statement that the state finds his behavior reprehensible, then the man will certainly suffer the distress of public humiliation and a decline in his reputation. And so, this final response maintains, denunciation is simply another form of punishment.

The problem with this final response, however, is that it, too, neglects the distinction between harming and intentionally harming an offender. If the state publishes a man's name in the newspaper in order to harm him, then it does indeed punish him. But if this is what the state does, then the argument from a right to censure the offender cannot justify it. If, on the other hand, the state publishes a man's name in the newspaper in order to publicly express its disapproval of his behavior, then the argument from such a right will justify it. But if this is the case, then any harm that befalls the offender will simply be a foreseeable consequence of the state's denunciation of him, so the state's act will not be a punishment. And so, in either case, the right to censure the offender, if there is such a right, will fail to justify a right to punish him and the reprobative solution will again fail the entailment test.

4.2.4 The Not Punishing the Guilty Objection

I have argued that the state does not have the right to censure in the strong sense of having a right required by the reprobative solution, and that even if it did have such a right, this would not suffice to justify punishing those who merit such censure. Even if I have been mistaken on both of these points, however, there is a final set of problems that undermines the reprobative solution. For even if the state does have the right to punish all who merit its censure, this is not the same as saying that it has the right to punish all (and only) those who break its just and reasonable laws. The reprobative solution, that is, will still prove subject to the problems of not punishing the guilty and of punishing the innocent and so will, once more, fail the entailment test (and, for many people, the foundational test as well).

The not punishing the guilty objection arises because a person can violates a just and reasonable law yet not merit disapproval. The person who knowingly drives his friend to the hospital in a car that has not had its required emissions inspection, for example, violates such a law but does nothing morally wrong. It is implausible to imagine a judge fining the driver by declaring that in doing so she expresses disapproval of his

behavior. Indeed, it is difficult to imagine her saying that she would have done anything different in those circumstances. But if punishment is justified, the judge surely has the right to enforce the fine in this case in particular and to impose punishment in cases where a person acts illegally but not immorally in general. As we saw in considering desert-based retributivism in Section 3.1.2, there are many cases in which a person violates a just and reasonable law but does not seem to merit moral criticism. A successful solution to the problem of punishment must justify the right to punish in such cases, but the reprobative solution clearly cannot do this. And so, as a result, the reprobative solution cannot provide a satisfactory solution to the problem of punishment.

4.2.5 The Punishing the Innocent Objection

The punishing the innocent objection arises because there are cases in which a person's acts remain within the limits set by just and reasonable laws but merit censure nonetheless. No one would think it objectionable, for example, if the state declared a certain day to be Holocaust Memorial Day. In doing so, the state would clearly be acknowledging that those who deny the facts of the Holocaust merit disapproval. It is difficult to accept that just and reasonable laws would criminalize the expression of their views, however. As a result, the state would have the right to censure people for doing things that its just and reasonable laws did not forbid. And the examples could be easily multiplied. It would be difficult to deny that people merit censure for telling offensive jokes, cheating on their boyfriends, girlfriends, or spouses, playing practical jokes at funerals, and so forth, and equally difficult to maintain that just and reasonable laws would criminalize such behavior. In all of these cases, then, the reprobative solution is saddled with the implication that the state may permissibly punish these people for acting in these ways (indeed, with the stronger claim that the state should punish them) even though they have broken no laws. This implication is likely to strike most people as so implausible that the solution cannot pass the foundational test. And it is clear, in any event, that the implication prevents the reprobative solution from drawing the line between those who may be punished and those who may not in the place where punishment draws it, and so prevents the solution from passing the entailment test. Either way, once more, the reprobative approach fails to provide a satisfactory solution to the problem of punishment.

4.3 THE MORAL EDUCATION SOLUTION

On December 14, 2001, Judge Alvin K. Hellerstein sentenced Edward Bello to ten months of home detention for conspiracy to use stolen credit cards, and in doing so he imposed a novel condition as part of his

sentence: Bello could not watch any television during that time. "It is . . . important that the normal diversion of television-watching be denied," Judge Hellerstein explained, "in order that he have ample opportunity to reflect on the ways of his life and the harm that he has brought to his family" [Weiser (2002: A24)].[17] In appealing to the claim that punishment can prompt an offender to reflect on the error of his ways, Judge Hellerstein's justification for this unusual sentence represents another alternative to the more familiar consequentialist and retributivist solutions to the problem of punishment. On this account, harming an offender is justified because harming him now will teach him a lesson that will benefit him later.

The claim that punishing an offender can ultimately benefit him in this manner has a long and venerable history. It should be familiar to anyone whose parents ever used the immortal words "I'm doing this for your own good." What I call the "moral education solution" uses this claim to solve the problem of punishment. As one of the most prominent defenders of this approach has put it, "by reflecting on the educative character of punishment we can provide a full and complete justification for it" [Hampton (1984: 113)].[18]

4.3.1 Punishment as Education

The moral education solution, like the reprobative solution, arises from the conjunction of two claims: the claim that the state should (or at least may) communicate the wrongness of the offender's behavior to the offender and the claim that if the state should (or may) do this, then it is permissible for it to do so by punishing the offender. Punishment, on this account, is a form of compulsory moral education. If the state does succeed in educating offenders about the wrongness of their behavior, of course, this is likely to reduce the chance that they will commit similar offenses in the future. And punishing people to deter them from committing future offenses is the hallmark of the consequentialist approach to

[17] The sentence was later stayed by a higher court and ultimately reversed.

[18] The moral education solution has also been defended by a number of other writers, including Duff (1986), Morris (1981) [although he treats it only as "one principal justification" for punishment: (46)], Gahringer (1960, 1969), Ewing (1929: chap. IV), Prust (1988) [although Prust characterizes the position he defends as a form of retributivism, it is clear that it is the same as what I am here calling the moral education solution, i.e., one that involves "defining the goal of punishment in terms of character change" of the offender (1988: 35); the same seems also to be true of Lemos (1977: 61ff.) and, at least on some interpretations, Plato, Hegel (e.g., McTaggart 1896), and Dewey (e.g., Shook 2004)]. A sympathetic account can also be found in Deigh (1984) and a defense of a closely related view, on which the punishment must be voluntarily accepted by the offender, in Khatchadourian (1982).

the problem of punishment. But the moral education solution treats this deterrent effect of punishment as merely a fortuitous side effect; the real aim is the moral improvement of the offender. As Hampton puts it, any deterrent effect produced by such education is "certainly welcome," but it is not the "goal" of punishment (1984: 117). The good to be promoted, on this account, is the offender's "moral growth," and the state thus "punishes him as a way of communicating a moral message to him" (119).[19]

Defenders of the moral education solution generally treat the first claim needed to sustain the position, the claim that the state should communicate the wrongness of the behavior to the offender, as uncontroversial. And, indeed, it may seem difficult to argue with the claim that moral enlightenment is a good thing.[20] The second claim, however, the claim that if the state should do this, then it may do so by punishing offenders, is far less obvious. Even if we agree that the state should teach offenders a lesson, we might wonder why this could not be accomplished by some other means, say by subjecting them to a stern lecture or to public censure. To this question, defenders of the moral education solution have offered two answers. For a message to be successfully conveyed by one person to another, they have argued, two things must happen: the person being addressed must pay attention to the person addressing him, and he must understand what this person is saying. One argument for punishment as the means of communication maintains that the harm caused by punishment is necessary to get the offender to pay attention to the message; another maintains that it is necessary to get him to understand the message itself.

The first argument treats the harm caused by punishment as a device for getting the offender's attention. As Primoratz puts it, for example,

[19] Strictly speaking, Hampton appeals to the value of educating the offender both to himself and to society as a whole: "the goal of punishment . . . is the offender's (as well as other potential offenders') realization of an action's wrongness" (1984: 117). In her later treatment of the moral education theory, Hampton says more explicitly that there are, in effect, two different forms that the theory can take: a "specific form" that "justifies punishment aimed at communicating the wrongness of his action to the criminal himself" and a "general form" that "justifies punishment as a way to educate the larger community about those values in the name of which the state is punishing" (1998: 40). The latter form is clearly a nonutilitarian version of the consequentialist solution, in which the offender is harmed as means of generating a benefit to others, and so has already been implicitly addressed and rejected in Section 2.4. The discussion in this section will therefore focus on the specific form of the moral education theory.

[20] For an objection to the claim that moral enlightenment really does benefit the offender, however, see Shafer-Landau (1991: 209–11). In addition, it is worth noting that one might question whether the state itself is an appropriate source for providing such enlightenment even if it is conceded to benefit the recipient. To the extent that people oppose the teaching of values in the public schools, for example, there might be a basis for rejecting this foundational claim.

"merely verbal condemnation is not likely to *reach* its immediate addressee. ... Regrettably, although perhaps not surprisingly, many criminals are *oblivious* to mere words" [1989b: 199, emphases added; see also (1989a: 152–3)]. Similarly, Gahringer says that "[p]unishment may serve as a language appropriate to occasions where words, *even though understood*, cannot break through the barrier of mere understanding to communicate with a self which wills to hold itself inaccessible" (1960: 47, emphases changed). On this account, a merely verbal explanation of the wrongness of the offender's behavior would suffice to make her understand it if only she pays attention to what is being said; however, precisely because she is an offender, she is unlikely to pay attention to a mere lecture. By contrast, she is very unlikely to ignore the fact that she is being harmed. Thus, punishment is a way of making sure that she does not simply ignore the message aimed at her.

The second argument for the claim that punishment is necessary for moral education appeals to the content of the lesson that we want the offender to learn: we want her to learn that her actions were wrong. Her actions were wrong because they wrongfully harmed someone. So, what we want the offender to understand is that her actions harmed someone and that being harmed is bad for people. And so, on this account, the harm we do to the offender is the language we use to communicate the message that her act wronged someone else. As Hampton puts it:

> [T]he person who wrongfully harms another is not thinking about the others' needs and interests, and most likely has little conception of, or is indifferent to, the pain her actions caused another to suffer. Hence, what the punisher needs to do is to communicate to the wrongdoer *that* her victims suffered and how much they suffered, so that the wrongdoer can appreciate the harmfulness of her action. How does one get this message across to a person insensitive to others? Should not such a person be made to endure an unpleasant experience designed, in some sense, to "represent" the pain suffered by her victim(s)? ... By giving a wrongdoer something like what she gave to others, you are trying to drive home to her just how painful and damaging her action was for her victims, and this experience will, one hopes, help the wrongdoer to *understand* the immorality of her action. (1984: 131, second emphasis added)

Even if an offender were willing to pay attention to the moral lesson being addressed to her, that is, mere words would be insufficient to convey the requisite meaning.[21]

[21] This claim has also been pressed by Duff in the context of an analogy with penance in certain religious traditions: "If the suffering to which [the sinner] is subjected is imposed on him in the right way and in the right spirit it can bring home to him, as merely verbal or symbolic condemnation would not, the extent and nature of his sin." (1986: 251).

Although the moral education solution to the problem of punishment is less common than the more orthodox consequentialist and retributivist positions, its central contention should be familiar. As Hampton points out, for example, it seems to be implicitly endorsed every time a parent punishes a child so that the child will "learn his lesson" or whenever a victim of a crime demands that the offender who wronged her be punished so that the offender will "understand" what he "did to me" (121–2, 120). There is little question, then, that the moral education solution offers a distinctive alternative to the consequentialist and retributivist approaches, and that it appeals to considerations that seem, at least on the face of it, to be well grounded. The question, then, is whether it is well grounded enough. And the answer, for several reasons, is no.[22]

4.3.2 The Not Punishing the Guilty Objection

The most common objection to the moral education solution concedes that the state should try to educate morally those who break the law but raises doubts about whether punishment is necessary to do so. Ten, for example, asks why a stern lecture could not suffice to get the message across (1990: 202–3), and Hershenov proposes the alternative of an official announcement of public censure (1999: 85).[23] Understood as a general attack on the moral education position, the appeal to such non-punitive alternatives is less than compelling. Primoratz's insistence that many offenders are "oblivious to mere words" seems plausible, and if this is so, then it can seem plausible that only punishment will drive the lesson home. But the objection can be pressed in a more focused manner. The critic can concede that in at least some cases, and perhaps even in many, nothing short of punishment will lead to the offender's moral enlightenment.[24] She need only maintain that this is not true in all cases. As long as offenders can be led to see the error of their ways without being punished, there will be some cases in which the moral education solution cannot justify the right to punish them. And since punishment involves treating the line between those who have broken the law and those who have not as morally relevant in determining how it is permissible to treat them, such cases will suffice to overturn the moral education solution by

[22] In addition to the objections raised in the text, it is worth noting briefly that the moral education solution runs counter to intuitions that most people seem to share: that the function of punishment is not to benefit the offender [see, e.g., Adler (2000: 1427)] and that if the state pardons an offender or declines to prosecute him, the state does not wrong him in any way [see, e.g., Husak (1992: 450)]. Some additional objections that also merit notice can be found in Golash (2005: chap. 6).

[23] This objection is also endorsed by Ten (1987: 45) and Braithwaite and Petit (1990: 163).

[24] Although see Marshall (1984) for an argument that punishment can never morally educate. Hobson responds to this argument in (Hobson, 1986).

demonstrating that it fails the entailment test. The question, then, is not whether punishment is ever needed to educate those who break the law, but whether there are ever cases in which it is not needed. The answer is that there are many such cases, and thus many good reasons to accept the not punishing the guilty objection as a reason to reject the moral education solution.[25]

4.3.2.1 *The Easily Reformed Offender Version*

One case involves offenders who do need something to prompt them to recognize the wrongness of their behavior but who can be made to see the light by something short of punishment. Even Primoratz insists only that "many" offenders are oblivious to mere words, and this leaves open the plausible corollary that many others are not. A first-time offender who has been convicted of a nonviolent offense such as trespassing or vandalism, for example, may need nothing more than a respectful interaction with someone who can make it clear to him how deeply upset the victim can be. If such interaction suffices to enlighten the offender, then the moral education solution cannot justify the state's right to punish him. In addition, many offenders can be made to see the wrongness of their behavior by being forced to interact with and assist, their victims. Indeed, Hampton herself at one point acknowledges not only that compulsory victim restitution can be an effective way of getting offenders to understand the wrongness of their behavior, but also that it is often likely to be more effective than punishment (1984: 132).[26] And an offender who was ignorant of the law she violated can be educated simply by clarifying the content of the law. For example a person who makes a right turn at a red light, mistakenly believing that doing so is legal in the state she is passing through, can be sufficiently educated by pointing out to her that such

[25] In addition to the cases discussed in what follows, it is worth noting that the moral education solution cannot account for the right to punish in cases where the offense is not morally wrong, such as the cases discussed in the context of desert-based retributivism in Section 3.1.2 in which, for example, someone drives a friend to the hospital in a car that has not passed a required emissions test.

[26] She suggests, for example, that a "youth charged with burglarizing and stealing money from a neighbor's house" might receive "the punishment of supervised compulsory service to this neighbor for a period of time" and argues that "such punishments seem to fit these crimes because they force the offender to compensate the victim." Hampton does not recognize that the concession that the moral education theory would only justify such sentences poses a problem for the theory itself because she mistakenly characterizes such treatment of the offender as a form of punishment. As we saw in Section 1.1.8, however, compulsory victim restitution is importantly different from punishment because the harm to the offender is foreseen rather than intended. Thus, in cases where compulsory victim restitution would suffice to morally educate an offender (and, like Hampton, I suspect that there would be many such cases), the moral education position cannot account for the claim that the state would have a right to punish the offender.

turns are illegal in that state. In all of these cases, then, there are nonpunitive means of morally educating lawbreakers. And so, in all of these cases, the moral education solution will fail to explain why the state has the right to punish someone who has broken the law. A defender of punishment is committed to the claim that the state does have the right to punish in these cases, so this failure of the moral education solution renders the solution unacceptable.

4.3.2.2 The Already Repentant Offender Version

The easily reformed offender provides one reason to accept the not punishing the guilty objection. An even stronger reason is provided by the offender who has already repented for his offense.[27] The moral education justification for punishing offenders is that punishment will teach them a valuable lesson. If a particular offender has already learned this lesson, then the moral education position can offer no justification for punishing him (or, indeed, for doing anything to him). Duff seems at one point to tacitly admit this in his defense of the reprobative solution, characterizing his position as one on which "the imposition of punishment on an *unrepentant* criminal is to be justified as a compulsory penance" (1986: 257, emphasis added), and the problem is just as dire for the moral education position. The concession, after all, implies that punishment is not justified when the offender has already learned his lesson. And one need not take an unduly sentimental view of people who break the law to recognize that there can be many such cases.

Many offenses, for example, are committed under the influence of drugs or alcohol. And in many such cases, it is plausible to suppose that the offender will fully grasp the wrongness of her behavior as soon as she becomes sober. A drunk driver who kills another person, for example, is unlikely to have difficulty appreciating the wrongness of his behavior once he sobers up. In addition, many people who are convicted of offenses involving negligence end up causing so much harm to people they care about that they are likely to suffer enough from the consequences of the offense to learn their lesson. A parent who unlawfully fails to put his child in a car seat, for example, will surely suffer enough to be repentant if the child dies or is seriously injured as a result of his negligence. And even if the child narrowly escapes serious injury, the close encounter itself will typically lead a parent to never make this mistake again. In all of these cases, too, therefore, the moral education solution will not justify the claim that the state has a right to punish people who have broken a just and reasonable law.

Hampton offers three responses to the problem of the repentant offender, but none of them are satisfactory. The first response is a

[27] See, e.g., Ten (1990: 204), Hoekema (1991: 342), and von Hirsch (1993: 10).

practical one: "Because it is difficult to be sure that a seemingly repentant criminal is *truly* repentant, and thus because a policy of suspending or shortening sentences for those who seem repentant to the authorities could easily lead the criminal to fake repentance before a court or a parole board, the moral education theorist would be very reluctant to endorse such a policy" (1984: 138). This response is unacceptable. It amounts to saying that we may punish truly repentant offenders because, if we make it a policy to not punish them, we will also inadvertently end up not punishing some nonrepentant offenders as well. But the objection maintains that on the moral education account, there is no right to punish genuinely repentant offenders in the first place. To say in response that we must punish those who do not merit punishment on educative grounds in order to punish all those who do merit punishment on such grounds is to justify harming those who do not merit punishment merely as a means of harming those who do. Such reasoning is incompatible with the moral education solution's rejection of the consequentialist approach to punishment, so it cannot be used to overcome the problem of the already repentant offender.

Hampton's second response to the already repentant offender objection appeals to the possibility that a particular offender may be repentant only because he is anticipating punishment. If this is the case, then even if he is repentant now, he won't be repentant once the state decides to forgo his punishment. And so, on this account, the state still has the right to punish a repentant offender to ensure that he remains repentant (138). The problem with this reply is simple. It succeeds in the case of offenders who become repentant only after they have been convicted, but it does not address people who become repentant as soon as they realize what they have done. The parent whose failure to put his child in a car seat leads to the child's death does not wait until he has been sentenced to feel terrible about what he has done. Indeed, if he is ignorant of the law in question, he does not even wait until he discovers that he has broken the law. And the same can be said of many other cases that motivate the already repentant offender objection. And so, once again, the moral education theory cannot justify the claim that a successful solution to the problem of punishment must justify in order to pass the entailment test: the claim that it is permissible for the state to punish these people.

Hampton's third response to the problem of the already repentant offender appeals to the plausible claim that "experiencing the pain of punishment can be a kind of catharsis for the criminal" (138). The problem here, it should be clear, is the same as the problem with the second response: while some offenders may need and receive such catharsis, others do not. The moral education theory cannot justify the claim that the state has the right to punish those who would not benefit from such cathartic suffering and so, once again, will fail to pass the entailment test.

Finally, it is important to note that although Hampton repeatedly attempts to fend off the unrepentant offender objection, in the end even she is forced to concede that "if there were clear evidence that a criminal was very remorseful for his action and had already experienced great pain because of his crime (had 'suffered enough'), [the moral education] theory would endorse a suspension of his sentence or else a pardon (*not* just a parole)" (138–9). But, as I have already noted, there are many such cases: well-meaning parents whose nonmalicious negligence tragically results in death or serious injury to their children, drunk drivers who kill their friends and neighbors, people whose trivial offenses unpredictably lead to serious harm to others, and so on. In all of these cases, Hampton must ultimately agree, the moral education solution cannot justify punishing people who break a just and reasonable law. Indeed, saying that such people should be granted a pardon puts things too weakly. The act of pardoning someone, like the act of forgiving him, entails that one has the right to do otherwise. But the problem is precisely that the moral education theory cannot show that the state has the right to punish such people in the first place. And this problem alone is sufficient to warrant rejecting the moral education theory as a solution to the problem of punishment. Punishment involves treating the fact that a person has broken a just and reasonable law as providing permission to treat him in ways that would otherwise be impermissible. But, as this concession makes clear, the moral education solution cannot do this. At most, it can justify treating lack of repentence for breaking the law as such a reason. And that is not enough.

4.3.3 The Punishing the Innocent Objection

As I have noted repeatedly, punishment draws a line between those who break the law and those who do not and treats that line as morally relevant, making it permissible to intentionally treat people on one side of the line in ways that we would never intentionally treat people on the other. A successful solution to the problem of punishment must therefore explain why it is permissible for us to draw this line in this way. I have argued that the moral education solution fails to do this for one important reason: it fails to show why everyone on the offender side of the line may permissibly be punished. But, like all of the other solutions to the problem of punishment we have considered, the moral education position also fails for a second reason: it entails that we may permissibly punish some people on the nonoffender side of the line. The moral education solution, that is, is also subject to the punishing the innocent objection.[28] This provides a second reason to conclude that the solution

[28] This problem has been noted by a number of writers, including Mabbot (1939: 39), Armstrong (1961: 153), Primoratz (1989a: 117), and Dolinko (1999: 357–8).

fails the entailment test and will also lead most people to conclude that it fails the foundational test as well.

The moral education solution has this problem for the simple reason that there are many people who do not break the law but who nonetheless need moral education: vicious racists whose behavior remains within the limits of the law, verbally abusive spouses, people who are unfaithful to their loved ones, who lie to their friends, talk behind people's backs, and so forth. In all of these cases, the person in question would benefit from moral education. And so, if it is true that punishing offenders is permissible because it benefits them by enlightening them, then it will also follow that punishing these people will be morally permissible for the same reason.

Hampton responds to the punishing the innocent objection by conceding that at least some people who are legally innocent "probably" do need moral education, but she maintains that this does not commit the moral education theorist to the conclusion that it would be permissible for the state to punish them:

> [T]he state should refrain from punishing immoral people who have nonetheless committed no illegal act, not because they don't need moral education but because the state is not the appropriate institution to effect that education. Indeed, one of the reasons we insist that the state operate by enacting laws is that doing so defines when it may coercively interfere in the lives of its citizens and when it may not; its legislation, in other words, defines the extent of its educative role .... So if the state were to interfere with its citizens' lives when they had not broken its laws, it would exceed its own legitimate role. (1984: 132–3)

And so, the moral education solution, on this account, justifies punishing people who need moral education, but only if they break the law.

This response to the punishing the innocent objection is unsuccessful for several reasons. First, it is widely agreed that criminal law does not define the extent of the state's educative role. Smoking tobacco, for example, is not illegal, but it is widely viewed as appropriate for the state to educate people about the dangers of smoking. Second, even if the acts that the state criminalizes limit the cases in which it is appropriate for the state to educate people, it does not follow that such education would be appropriate only when a person engages in those acts. Even if only the fact that rape is illegal makes it appropriate for the state to interfere with rapists, for example, it does not follow that it is appropriate for the state to interfere only with those who actually rape. A man who publicly states that all women deserve to be raped, for example, or that it is morally permissible for men to rape their wives, may break no law, but the state's presumably legitimate interest in preventing rape would seem to give it as good a reason to educate him as to educate those who commit rape.

Finally, and most fundamentally, Hampton's response to the punishing the innocent objection begs the question. Her response depends on the claim that when the state passes a law prohibiting a given behavior, this makes it "legitimate" for the state to "coercively interfere" in the lives of those who nonetheless engage in that behavior. But whether or not the state's prohibition of a given behavior renders such a response morally permissible is precisely the question that the moral education theory is attempting to answer. The theory attempts to answer the question in the affirmative by pointing to the educative value of punishing people for breaking the law. But if this is the feature of punishing lawbreakers that renders their punishment permissible, then the moral education theory cannot avoid the implication that punishing immoral law-abiding people would also be permissible, since the feature that justifies punishment in the former case is also present in the latter. If, on the other hand, the state has not passed a law prohibiting a given behavior, and therefore may not punish people for engaging in it, then the reason the state may punish those who engage in unlawful behavior must be that the state has made such punishment legitimate by prohibiting that behavior. This would be to retreat to the logical entailment argument in favor of punishment that was considered and rejected in Section 1.2.3 rather than to salvage the moral education solution that Hampton and others have attempted to defend.

4.3.4 The Paternalism Objection

Defenders of the moral education solution often note that parents typically invoke the educative function of punishment when they justify punishing their children [e.g., Gahringer (1960: 46–7); Hampton (1984: 120–2)]. And, indeed, the fact that a parent often pictures punishment of a misbehaving child as a way of "teaching him a lesson" – a harsh treatment that is done "for his own good" – does give the moral education position an initial surface appeal. As Hampton puts it, such considerations render the moral education theory "intuitively very natural and attractive" (120). But the analogy with parental punishment in the end also generates a serious additional difficulty for the moral education solution. For we beleive that it is appropriate for parents to have a paternalistic relationship with their children but inappropriate for the state to have such a relationship with its adult citizens.[29] The moral education theory maintains that it is permissible to harm an offender now because doing so will benefit her later. But in other relevantly similar situations, we would not consider it permissible for the state to harm a

[29] The paternalism objection has been pressed by Murphy (1985: 25–7) and Golash (2005: 123–6).

citizen now merely because this will benefit her later. The paternalistic foundation required by the moral education solution, that is, has implications that most people find morally unacceptable. And so, even if the solution could pass the entailment test by justifying the permissibility of punishing all and only those who break the law, it would still seem incapable of passing a reasonable application of the foundational test.

Consider, for example, a seriously overweight person. In the short run, being forced to diet and exercise will harm him, but in the long run it will benefit him. Or consider pulling a person's tooth, which will cause a lot of pain now but prevent even more pain later. In these cases, it would benefit the person to be subject to painful treatment now. But in such cases, we would surely object if the state compelled the person to undergo the treatment and attempted to justify this practice by appealing to the long-term benefits the person would receive. And yet, on the moral education solution, forcing an unenlightened offender to suffer now so that he can be morally healthy later is no different from forcing an overweight person to diet and exercise now so that he will be physically healthy later.[30]

Hampton acknowledges that the moral education solution may at first seem objectionably paternalistic, but she insists that this appearance is deceptive:

> [W]hen such philosophers as John Stuart Mill have rejected paternalism, what they have rejected is a certain position on what should be law; specifically, they have rejected the state's passing any law which would restrict what an individual can do to *himself* (as opposed to what he can do to another). They have not objected to the idea that when the state justifiably interferes in someone's life *after* he has broken a law (which prohibited harm to another), it should intend good rather than evil towards the criminal." (1984: 122–3)

And so, on Hampton's account, the moral education solution can evade the charge of paternalism: it would be wrong for the law to prevent people from overeating for their own good, because that would paternalistically prevent people from harming themselves, but it is not wrong for the state to punish people for their own good, because that would not prevent people from harming themselves, paternalistically or otherwise.

Hampton's response to the paternalism objection is unsatisfactory because it begs the question. The response maintains that it is permissible for the state to interfere in a person's life for his own good when it is justifiable (as in punishing an offender for his own good) but not

[30] This objection is also effectively pressed by Dolinko, who notes that we would not consider it permissible for the state to compel adults to undergo medical treatment without their consent even if the treatment were needed to save their lives (1999: 354–5).

permissible for the state to interfere in a person's life for his own good when it is not justifiable (as in compelling an overweight person to diet and exercise). But whether or not it is permissible for the state to interfere in the offender's life but not the overweight person's life is the question in the first place. If the benefit to the offender makes it permissible to punish him, then the benefit to the overweight person also makes it permissible to compel him to diet and exercise. If the benefit to the offender is a legitimate consideration only because harming him is already permissible for some other reason, then that other reason, and not the benefit to him, explains why it is permissible to punish him. On the former alternative, the defender of punishment cannot escape the paternalism objection; on the latter, he cannot depend on the moral education solution. And so, on either account, the moral education approach once again fails to provide a satisfactory solution to the problem of punishment.[31]

4.4 THE SELF-DEFENSE SOLUTION

I have argued that backward-looking and forward-looking considerations cannot provide a satisfactory solution to the problem of punishment and that appeals to other considerations are unsuccessful as well. If I am correct, then there would seem to be only one final option for the proponent of punishment: an attempt to solve the problem of punishment by appealing to a combination of some of these considerations. The question then becomes how to go about producing such a combination. One answer appeals to a specific context in which a particular combination of backward- and forward-looking considerations does seem to justify the infliction of harm on other people: the practice of individual self-defense. I will consider the permissibility of punishment on grounds of self-defense in this section. The other answer turns on a more general thought about how individual moral theories can be combined into hybrid theories. I will consider this approach in the section that follows. I will argue that both of these final attempts to solve the problem of punishment also fail and that we have, therefore, no successful solution to the problem of punishment.

[31] It is also worth noting that Morris offers a contractarian alternative to Hampton's reply to the paternalism objection: since each of us knows that we are a potential evildoer, Morris argues each of us has reason to put in place a system that would promote our chances of being morally good people (1981: 50). But this argument, too, is incapable of distinguishing punishment from other clearly objectionable forms of paternalism. After all, it could equally be noted that each of us knows that we have the potential to become seriously overweight, and so would favor a system that would promote our chances of being healthy people.

4.4.1 Punishment and Self-Defense

So suppose, to begin with the self-defense solution, that Larry fires a gun at Moe, and the only way for Moe to avoid a fatal injury is to fire back at Larry. Virtually everyone will agree that it is morally permissible for Moe to shoot at Larry under such circumstances. But neither a purely forward-looking nor a purely backward-looking analysis provides an adequate account of this judgment. The justification for the claim that it is permissible for Moe to shoot Larry in self-defense lies, at least in part, in forward-looking considerations. Moe's shooting Larry now will produce a positive consequence in the future: it will save Moe's life. And this fact clearly plays an important role in rendering Moe's act morally permissible. If shooting at Larry now would not produce this positive consequence (if, for example, Moe had already suffered an obviously fatal injury), then shooting at Larry now could not be justified as an act of self-defense. But the justification for Moe's right to shoot at Larry now also lies partly in backward-looking considerations. It is because of Larry's past wrongful attack on Moe that the fact that shooting Larry now will save Moe's life justifies shooting him. If Larry's past act of shooting at Moe had not been wrongful in the first place (if, for example, Larry began shooting at Moe only because Moe himself had first started shooting at Larry and because shooting at Moe was now the only way for Larry to protect himself from Moe's wrongful attack), then Moe's shooting at Larry now could not be morally justified as self-defense even if it was necessary to save Moe's life.

The claim that it is morally permissible for individuals to harm others in self-defense is relatively uncontroversial. And the justification for the claim seems clearly to rest on a combination of backward- and forward-looking considerations. Self-defense therefore seems to provide an attractive model for an attempt to solve the problem of punishment by using a similar combination. In addition, the social practice of punishment might at first seem to be nothing more than the individual practice of self-defense writ large. Why do we seek to capture and punish those who break the law, after all, if not to protect ourselves from them? Given all of this, it might seem that a defense of punishment by appeal to the right to self-defense would be a relatively simple matter. But this appearance is deceptive. For there is a crucial difference between harming a person in self-defense and harming a person as punishment. When Moe harms Larry in self-defense, he harms Larry in order to *prevent* Larry from wrongfully harming him. But when the state punishes an offender, it punishes him precisely because he has already *succeeded* in wrongfully harming someone. It is easy to see how the notion of self-defense can justify harm to prevent a particular wrong from taking place. But it is far more difficult to see how an appeal to self-defense could be used to justify inflicting harm in response to a particular wrong when it is

already too late to prevent that wrong from taking place and thus too late to provide a defense against it. Several writers have nonetheless argued that a self-defense solution to the problem of punishment can overcome this difficulty. I will argue in this section that they are mistaken.[32]

4.4.1.1 The Right to Threaten and the Right to Punish

Punishment takes place after a wrong has been committed; self-defense prevents a wrong from occurring. This poses a problem for any attempt to justify punishment on grounds of self-defense. The attempt to overcome this problem turns on the distinction between the right to *inflict* punishment, on the one hand, and the right to *threaten* to inflict punishment, on the other. On most accounts of punishment, the latter right is parasitic on the former. We have the right to threaten to punish people for breaking the law, that is, because we have the more fundamental right to actually punish people for breaking the law (and because we have the right to threaten to do whatever we have the right to do). If this is the correct account of the relation between the two rights, then the self-defense solution is doomed. We punish people after they commit offenses, not before, so we cannot justify the right to punish people by appealing to the right to attempt to prevent them from committing offenses. But the self-defense solution can still proceed if it maintains that the relation between these two rights is precisely the opposite: that it is the right to threaten to punish people for breaking the law that is fundamental, and from this basic right we can then derive the further right to punish people for breaking the law (since we have the right to do whatever we have the right to threaten to do). On this account of the relation between the two rights, the self-defense solution may be able to account for the fact that punishment follows an offense, while self-defense involves preventing one: we cannot punish a person to prevent him from committing an offense that he has already committed, but we can *threaten* ahead of time to punish him if he commits the offense, and we can issue the threat to try to prevent him from committing it. Self-defense, that is, may turn out to justify the right to threaten to punish. And if that right can, in turn, justify the right to punish, then self-defense may in the end justify the right to punish after all. This defense will therefore ultimately rest on two claims: the claim that it is morally permissible for the state to threaten to punish people for breaking the law and the claim that if this is morally permissible, then it is morally permissible for the state to punish people who break the law despite being threatened.

[32] In addition to the defenses of the self-defense solution treated in this section, see Hurka (1982) and Ellis (2003). In addition to the objections to the solution raised in this section, see also Golash's criticisms of the self-defense solution in general (2005: chap. 5), and McKerlie's (1983) rebuttal to Hurka (1982) in particular.

4.4.1.2 The Right to Threaten to Punish

The most basic claim needed to sustain the self-defense solution is the claim that we have the right to threaten to punish people for breaking the law. Two arguments have been offered in defense of this claim. One argument is made by Quinn, who attempts to justify the claim by appealing to an analogy with what he calls "m-punishment" (1985: 57ff.). Suppose, to summarize his example, that there were mechanical devices that, once activated, would automatically identify and capture those who break the law and subject them to precisely the treatments that the state subjects convicted offenders to when it punishes them. And suppose further that these devices would do this with the same efficiency and accuracy that the state achieves when it does these things. Quinn argues that it would be morally permissible for the state to publicly activate such devices as a deterrent threat, and that activating such devices would be morally equivalent to establishing a legal system threatening to punish those who break the law. His defense of the claim that the state has the right to threaten to punish people for breaking the law, therefore, rests on the conjunction of two claims: the claim that it would be morally permissible for the state to activate the m-punishment devices and the claim that activating these devices would be morally equivalent to threatening to punish people for breaking the law.

Quinn's argument for the first of these claims, the claim that it would be morally permissible to activate the m-punishment devices, in turn, rests on a series of analogies. Suppose, to use a specific example, that Larry has a house and he fears that Moe will attempt to break into it, perhaps to steal Larry's property or to harm Larry. Then surely, the argument begins, self-defense would justify the permissibility of Larry's constructing a fence around his house to deter Moe from attempting to break in. But suppose, further, that a simple fence would be unlikely to deter Moe. In that case, the argument continues, it would surely be permissible for Larry to put sharp spikes on the outside of the fence so that Moe would incur a significantly higher cost by breaking into Larry's home. If self-defense renders construction of the fence permissible, that is, then it must render construction of the spikes permissible as well. But suppose, in addition, that for some reason it is not possible to place spikes outside the fence but it is possible to place them on the inside, and in such a manner that Moe can clearly see them from the outside. This third fence, the argument maintains, is morally no different from the second. In each case, a fence is built so that Moe will see that if he breaks into Larry's house, he will incur a significant cost. In the former case, it is true, he will incur the cost on the way in, while in the latter he will incur it on the way out, but this fact in itself is not, the argument claims, morally relevant. What matters is that in both cases, the fence is put there to give Moe a good reason not to try to break in. And so, if the outer-spiked fence is permissible, then so is the inner-spiked fence.

Once we concede that construction of the inner-spiked fence is permissible, the remainder of the argument seems to follow fairly straightforwardly. For if Larry can construct a fence with spikes on the inside, then surely he can construct a fence with some other device on the inside designed to inflict the same amount of harm. So he could, for example, place an automated device inside the fence that would subject Moe to the same harm that the inner-spiked fence would have imposed. And, finally, if Larry could program what Quinn refers to as an "autoretaliator," in which the amount of harm it would inflict on Larry could be made to vary, depending on the wrong that Moe committed (e.g., so that it would inflict more harm on Moe if he had killed Larry than it would if he had stolen Larry's computer), then this, too, would be permissible as a means of protecting Larry from Moe. This last conclusion amounts to saying that it would be permissible to activate what Quinn refers to as an "m-punishment device," and since doing this would, he maintains, be "functionally equivalent" to setting up a justice system that threatened to punish people for breaking the law, it follows from all of this that it is morally permissible for the state to threaten to punish people for breaking the law.

A second and very different argument for the claim that the state has the right to threaten to punish people for breaking the law appeals to principles of distributive justice. This argument has been most prominently defended by Montague (1995, 2002) but also by several other writers, including Farrell (1990) and Cederblom (1995).[33] The argument begins with a principle of distributive justice that is meant to govern cases of a certain sort: if there is a situation in which it is inevitable that either *A* or *B* will be harmed, and if this situation is *A*'s fault, then it is just to distribute the harm to *A* rather than to *B*. The point is not the desert-based claim that it is good for *A* to suffer in such cases. It is the considerably more modest claim that since someone is going to be harmed, it is better for *A* to suffer rather than *B*. Suppose, for example, that Larry wrongfully tosses a rock in the air, that the rock is going to fall on Moe's head, and that the only way to prevent this would be to deflect it so that it ends up landing on Larry's head. In this case, it is inevitable that the rock will fall either on Larry or on Moe, and it is Larry's fault that this is so. Given these facts, the principle maintains that it would be just to deflect the harm to Larry, not because it is good that Larry suffer for his wrongful act, but simply because it is less bad (or at least less unfair) that he suffer rather than Moe.

To this normative claim, the proponent of the distributive justice argument then adds a descriptive claim about society: either the state threatens to harm people for breaking the law or it does not. If it does

[33] Cederblom refers to the position he defends as the "retributive liability theory of punishment" and says that it "can be classified broadly as retributive" in its approach to punishment, but in substance his position is the same as that of Montague and Farrell.

not, then many innocent people will be harmed. If it does, then many innocent people will be spared from harm, but those who persist in breaking the law will be harmed. So, regardless of what the state does, someone will be harmed. The only question is whether it will be offenders or nonoffenders. But it is the fault of the would-be offenders that this is so. If no one were disposed to break the law in the first place, then it would not be inevitable that anyone will be harmed. From the conjunction of the normative principle of distributive justice and this descriptive claim about the choice that the state faces, it follows that it would be just for the state to threaten to punish people for breaking the law. Since it is the fault of offenders that someone must be harmed, it is just to make them suffer rather than innocent people. As Cederblom puts it, the justification for the deterrent threat on this account "is not to minimize harm, but to distribute it toward the guilty rather than toward the innocent" (1995: 307).[34]

4.4.1.3 The Right to Carry out the Threat to Punish

The proponent of the self-defense solution thus has two distinct ways to vindicate the solution's first claim: the claim that the state has the right to threaten to punish people for breaking the law. If this claim can be sustained, the defender of the solution must then justify the second claim: the conditional claim that if the state has the right to make this threat, then it also has the right to carry it out. Two arguments have been offered for this claim as well.

One argument turns on a claim about the relationship between threats and intentions. Quinn, for example, argues that if the state has the right to threaten to punish, then it has the right to form a conditional intention that it will punish an offender if he breaks the law. But it can then seem plausible to insist that if someone has the right to form an intention to do something, then he has the right to act on that intention. And so, if we grant that the state does have the right to threaten to punish people for breaking the law, we must concede that it has the right to carry out that threat.[35]

[34] See also Farrell (1988: 443; 1990: 302–3; 1995: 224) and Montague (1995: 62–3, 77).

[35] Although this argument is suggested by Quinn's remarks here, it is not entirely clear how much weight he wishes to place on it. He also seems to suggest that we should accept the claim simply because, if we accept it, we will end up with a defense of punishment that is much more attractive than the consequentialist and retributivist alternatives. Since this would clearly beg the question in favor the view that punishment is permissible, I focus here on the argument that Quinn seems to offer in defense of the claim, even if his defense is somewhat tentative. Otsuka (1996; 2003: 57–65) defends a modified version of Quinn's position on which it is permissible to carry out the threat to punish someone on the grounds that following through on a threat made to one person will deter others from aggressing when they are comparably threatened. This variant of the self-defense solution, however, is ultimately a version of the consequentialist solution, on which

Montague offers an alternative justification for the second claim, one that he maintains "omits Quinn's dubious claims about intentions" (1995: 72). He presents the central idea as follows:

> [I]f establishing a system of punishment is justified, then creating certain positions that system requires, with their attendant responsibilities, is justified, and the people who occupy those positions are certainly justified in fulfilling their responsibilities – that is, presumptively justified in participating in the punishment of individuals .... There is a clear sense, then, in which the justification of a system of punishment with the threats it embodies is sufficient to create presumptions in favor of punishing individuals within the system. (72)

On either account, the right to threaten to punish would entail the right actually to punish and thus, when combined with the claim that the state does have the right to threaten to punish, would suffice to vindicate the self-defense solution.

4.4.2 The Punishing the Innocent Objection

The self-defense solution is perhaps the most innovative and promising of the various nonstandard solutions that have been offered to the problem of punishment. In the end, however, it, too, is subject to several decisive objections. The first objection involves cases that involve threatening to punish innocent people. The problem is perhaps most clearly approached by beginning with the first claim made by the self-defense solution, with Quinn's autoretaliator argument in its defense.[36] So, to return to the case in which Larry wishes to build a protective fence around his property, suppose that, for some reason, Larry is unable to attach jagged spikes on the outside of his fence but is able to coat the fence with a visible, moderately toxic substance. Any prospective intruder would see that if he attempted to scale the fence, he would come into contact with the substance and would incur an amount of harm comparable to that produced by the spiked fence. Clearly, if Larry is allowed to build the spiked fence, he is allowed to build the toxic fence as well. But now suppose that the prospect of becoming sick is not enough to deter most intruders from climbing over the fence. And suppose also that a

punishment is permissible because of its beneficial consequences where the consequences are measured in terms of their promotion of self-protection. And it must therefore be rejected for the various reasons that were given for rejecting other versions of the consequentialist solution in Chapter 2. If it is permissible to punish an aggressor whom you have previously threatened to punish because this will increase the deterrent effect of your future threats to others, for example, then it is permissible to frame and punish an innocent person whom you have previously threatened for precisely the same reason.

[36] For a different way of developing this objection to Quinn, see Carcasole (2000: 231).

second toxic substance is available that would also make an intruder sick but would in addition have the following property: it would stick to the intruder's body for a long time and eventually infect all of his children. And suppose, finally, that although many would-be intruders would not be deterred by the prospect of being made sick by climbing the fence, they would be deterred by the prospect of infecting their children and causing them to suffer. In this case, it seems clear that the appeal to the right to deter would-be intruders that justifies coating the fence with the first toxin would also justify coating it with the second. But coating the fence with the second toxin amounts to activating a primitive m-punishment device by which the builder threatens to punish not only offenders but their innocent children as well. The threat, to be clear, is not directed to the would-be intruder's children. The threat itself is directed to the would-be intruder, since he is the one that the threat aims to deter. But the content of the threat nonetheless includes prospective harm to the children of anyone who scales the fence. If the autoretaliator argument justifies threatening to punish guilty people for breaking the law, therefore, it also justifies threatening to punish their innocent children as well. And since this latter implication is clearly unacceptable, so is the autoretaliator argument for the first claim made by the self-defense solution.

Before moving on to consider whether the distributive justice argument can avoid this problem of punishing innocent people, let me first note a few ways in which a defender of the autoretaliator argument might attempt to reply to the objection. There would seem to be only three possibilities. One would be to deny that under this argument it would also be permissible to threaten to harm the innocent children of people who break the law. But there seem to be no grounds for the proponent of the self-defense solution to sustain this claim. One could, of course, argue that it is wrong to threaten to harm such children because this involves threatening to do something that would be independently wrong (say, on the grounds that they do not deserve to suffer or that harming them would be unfair), but this would amount to simply abandoning the self-defense solution in favor of a desert-based or fairness-based solution rather than to vindicating the self-defense solution itself.[37] The point of the self-defense solution, after all, is to try to justify the right to make

[37] Quinn himself does nothing to provide a basis for such a response. At one point, in describing the m-punishment devices, he stipulates that they can be programmed to avoid punishing innocent people and says that "as it happens, we prefer" to program them in this way (1985: 58). But he gives no reason for this preference, and the worry is not that his position implies that we should prefer to punish the innocent but that it implies that it would be permissible for us to do so if we prefer to do so.

certain threats without having to first establish the independent permissibility of doing the threatened acts.

A second response to the punishing the innocent objection would be to concede that under the autoretaliator argument, it would be morally permissible for the state to threaten to harm innocent people, but to maintain that it would be impermissible for the state to actually carry out this threat. This is the view that many people hold, for example, about the morality of nuclear deterrence. They believe that one state has the right to credibly *threaten* to kill millions of innocent civilians in another state with a nuclear counterattack as a means of deterring that state from launching a nuclear first strike against it, but they also believe that if the threat failed to deter and the state was attacked, it would be horribly wrong to actually carry out that threat by retaliating, since this would kill millions of innocent people and would be too late to prevent the first strike from occurring. This view may prove to be a promising one to hold on its own terms, but it is unavailable to a defender of the self-defense solution. For while this response would save the solution's first claim from an unacceptable implication, it would do so only by undermining the solution's second claim. The claim that the state has a right to threaten to punish the guilty, that is, would no longer suffice to justify punishing people, because the right to threaten would no longer entail the right to carry out the threat.

Finally, a defender of the self-defense solution could bite the bullet and agree that her position entails that it would be permissible to punish innocent people. But as we have already seen in many previous contexts, this, too, is an unacceptable response. Punishment draws a line between those who break the law and those who do not and treats that line as morally relevant. The problem of punishment is the problem of accounting for its permissibility. Any solution subject to the punishing the innocent objection cannot draw the line between those we may intentionally harm and those we may not in the way that punishment demands, and for that reason alone, it cannot pass the entailment test required for any successful solution. In addition, for most people, the implication that innocent people may permissibly be punished in such circumstances is sufficiently counterintuitive to believe that the solution cannot pass a reasonable application of the foundational test as well. And so, the problem of punishing the innocent is a decisive problem for the self-defense solution, at least when its crucial first claim is defended by appealing to the autoretaliator argument favored by Quinn.

A defender of the self-defense solution can concede this objection and turn instead to the distributive justice defense of the solution's first claim. But this approach in the end fares no better. Suppose, after all, that the state finds itself in the same position faced by the fence builder who is choosing between the two different toxins with which to coat his fence.

If the state threatens to do nothing at all in response to violations of the law, many innocent people will be harmed. If it threatens only to harm offenders, this will deter some people from breaking the law but not others. And if it threatens to harm offenders and their children, this will deter even more people from breaking the law. Suppose, more specifically, that if the state chooses this third option, some people will still break the law, but for every one innocent child who will end up being harmed as a result, five other innocent people (maybe even innocent children) will be saved from being victimized by offenders in the first place. When faced with this choice, the distributive justice approach implies that the state is entitled to shift the harm as much as possible away from the innocent and toward the guilty. Threatening to harm offenders themselves will help accomplish this. But threatening to harm offenders and their children will accomplish even more. This solution will do the best job of shifting harm from the innocent to the guilty and so will be endorsed by the distributive justice approach.[38] But, once again, it will produce unacceptable results. Regardless of how the proponent of the self-defense solution defends the claim that the state has the right to threaten to punish offenders, it will also entail that it has the right to threaten to punish innocents as well. Either this will mean that the state also has the right to punish innocents, which will render the solution unacceptable, or it will mean that the right to threaten does not entail the right to punish, which will also render the solution unacceptable. So, either way, the self-defense solution cannot escape the problem of punishing innocent people.

4.4.3 The Disproportionate Punishment Objection

The first problem with the self-defense solution arises when we consider its implications for the treatment of the legally innocent. Two further problems arise when we consider its implications in certain cases involving the legally guilty. The first such problem concerns the amount of punishment that the solution will justify inflicting on those who break the law. It maintains that, at least in certain cases, the permitted punishment will not be proportionate to the offense. As in the case where

[38] A defender of the distributive justice approach might object that his principle applies to cases in which harm can be shifted from an innocent person to a guilty person but not to cases in which harm can be shifted from a larger number of innocent people to a smaller number of innocent people. But it is difficult to imagine how such a restriction could be motivated without producing a further set of even more unacceptable results. In the famous trolley problem, for example, virtually everyone agrees that it would be morally permissible to pull the switch on a runaway trolley so that it ended up killing only one innocent person rather than five, but such a restricted view of distributive justice would be unable to account for this.

this disproportionate punishment objection has been raised against other solutions, a proponent of the self-defense solution can bite the bullet and agree that this account justifies punishment that, intuitively, is severely disproportionate. Accepting this implication is consistent with the claim that the solution nonetheless justifies punishment and so is consistent with the claim that the solution passes the entailment test. But most proponents of the solution, such as Quinn and Farrell, are unwilling to do this, and so they, along with virtually everyone who is not already committed to the self-defense position, take this implication as a further reason to reject the attempt to ground the right to punish in the right to self-defense. Even if it cannot show that the solution fails the entailment test, it prevents it from passing a reasonable application of the foundational test.

The disproportionate punishment objection can be generated in two ways. In the too much punishment version, the self-defense solution justifies the right to inflict severe punishment for trivial offenses. Alexander, for example, raises this concern in the context of thieves who continue to try to steal his roses (1991: 324–5). He believes that it would be permissible for him to move his rose bushes to a private island surrounded by shark-infested waters (provided that, as in the cases appealed to by Quinn and others, the prospective hazards are made clear to the would-be offenders ahead of time), and that if this is so, then it would be equally permissible for him to construct a functionally equivalent moat to protect his rose bushes. It is difficult to see how a defender of the self-defense position could deny such claims. But it is equally difficult to see how, once such claims are accepted, the self-defense account can avoid the conclusion that it would be morally permissible for the state to inflict a painful death on those who steal roses. If it is permissible to build a lethal moat to protect one's roses, then, according to the argument that attempts to justify the self-defense solution, it must be permissible to threaten lethal consequences for stealing roses and, finally, to inflict those consequences on those who steal them. But a painful execution for petty theft will strike virtually everyone as morally unacceptable. And so, for virtually everyone, this version of the disproportionate punishment objection should provide sufficient reason to conclude that the self-defense solution fails the foundational test.[39]

The too little punishment version of the disproportionate punishment objection can be generated by considering the case of arson. Suppose that the threat of a $500 fine would be enough to deter the vast majority of

[39] For an earlier version of Alexander's argument, with different examples, see also Alexander (1980: e.g., 209–10). It should be noted, however, that although Alexander convincingly argues that the self-defense solution has this implication, he does not consider this fact to be a decisive objection to the position, but simply an important feature of it (1991: 328). The same problem is also pressed by Carcasole (2000: 231–2).

people from burning down someone's house. Suppose further that those who would not be deterred by this threat desire to commit arson so strongly that nothing short of a threat of death or torture would deter them. In this case, to protect our houses from being destroyed by arsonists, the self-defense solution would permit us to threaten them with a $500 fine, but it would not permit us to threaten them with death or torture (and even if it did, that itself would be a problem). More importantly, it would not, for example, entitle us to threaten a greater fine or a five-year prison sentence, because we are only entitled to threaten to use the least force necessary to protect ourselves, and increasing the fine or adding a jail sentence would provide no further protection. But since the right to punish is derived only from our right to threaten to punish, this means that since we would have the right only to threaten a $500 fine against arsonists, we would have the right to punish them only in that amount. To many people this will seem an unacceptably light sentence, again suggesting that the self-defense solution cannot pass the foundational test.

4.4.4 The No Excuses Objection

Cases like the arson example just described may prove to be uncommon in practice. This fact does not undermine the force of the objection, which points to the fact that the self-defense solution has no principled way of assuring proportional punishment. This is an objectionable feature of the solution, and it is objectionable regardless of how pervasive its ill effects would be in practice. There is a second class of offenses that raises a second problem of proportionality, however, and this class is all too common: cases in which an offender has a mitigating excuse. Like the problem of punishing the innocent, the problem posed by mitigating excuses is perhaps most easily framed in terms of the autoretaliator argument. But, also like that objection, it can be used to undermine the distributive justice argument as well.

Let us return to the case of Larry the fence builder. This time, Larry is constructing a portable fence to deter people from physically assaulting him. For some reason, Larry is unable to attach harmful devices to the outside of the fence but is able to attach a wide variety of such devices to the inside and to make them clearly visible to would-be attackers on the outside. After experimenting with a variety of such devices, he settles on a level of harm the threat of which seems to deter most people from attacking him. But then Larry realizes that he has calibrated his primitive m-punishment device using calculations designed to protect him from unprovoked attacks. Larry knows himself pretty well. He recognizes that he gets a bit carried away at times and often ends up provoking people with his inconsiderate and at times downright rude and offensive (but legal) words and actions. So he begins to wonder how useful his fence will be in

protecting him from attacks in those cases in which his behavior, although well within his rights and perfectly lawful, nonetheless infuriates people.

There are two ways of imagining how an autoretaliator device might be used in response to this problem, but neither one can give the self-defense solution an intuitively acceptable position on the relevance of mitigating excuses. The first possibility is that Larry's fence cannot be made to distinguish between provoked and unprovoked attackers. If this is the case, then the fence can only be put at one "setting," threatening the same degree of harm to every attacker, provoked or unprovoked. But this means that any defense of punishment grounded in Larry's right to make a conditional threat in self-defense will entail that we have the right to aim just as much punishment at provoked attackers as we do against unprovoked attackers. On the self-defense account of punishment, that is, there can be no mitigating excuses. And this implication will surely strike most people as sufficiently objectionable to warrant concluding that the solution fails the foundational test.

The second possibility is that Larry's fence can be made to distinguish between provoked and unprovoked attackers, and can be programmed to threaten to cause different kinds of harm in the two cases. At first, this might seem to make things better for the proponent of the self-defense solution. But, in fact, it makes things worse. For whatever level of harm Larry threatens to deter most unprovoked attackers, he will have to threaten a *greater* amount of harm to deter most provoked attackers. When people are provoked, after all, they are less likely to consider in a calm, impartial manner the long-term negative consequences of their behavior and are more likely to focus on the presumed short-term satisfaction they would get from attacking their provoker. If the prospect of harm comparable to that caused by spending three years in prison is enough to deter most unprovoked attackers, for example, then Larry will have to threaten harm comparable to something more like five years in prison to deter most provoked attackers. And this will mean that under any defense of punishment grounded in Larry's right to make a conditional threat in self-defense, we have the right to aim *more* punishment at provoked attackers than at unprovoked attackers. Not only can there be no mitigating excuses, that is, but what should count as mitigating excuses must instead count as aggravating circumstances, making the appropriate amount of punishment more rather than less. And this is surely an even more unacceptable implication.

The autoretaliator argument, therefore, is subject to the no excuses objection, and for most people this will clearly be a sufficient reason to reject it. What about the distributive justice argument? This argument is grounded in the claim that when harm will inevitably befall someone, it is fair to ensure that it falls on those who are responsible for this fact. In the case of physical assault in particular, the argument maintains either that law-abiding citizens will be harmed by unlawful attackers or that those

who would commit such attacks will be threatened with a certain amount of harm if they commit them. At this point, the argument can then be developed in two different directions, which correspond to the two possibilities identified for the autoretaliator argument. And again, as in the autoretaliator argument, neither possibility provides a satisfactory position for the self-defense solution.

The first possibility is to treat all cases of physical assault, provoked or unprovoked, as falling into a single class. On this approach, the appropriate amount of harm with which to threaten would-be attackers would be determined by what threat would, on the whole, do the best job of shifting harm to those who are responsible for the problem. The amount of harm so determined would then be applied to all cases of assault, provoked or not. As a result, the self-defense solution would, on this approach, entail that the punishment for a provoked or an unprovoked attack would be the same. And this result, as we have already seen, will strike most people as unacceptable.

The alternative, again, would be to divide the cases into two classes, provoked and unprovoked assault, and to determine the appropriate punishment for each separately. This possibility might, again, at first seem to be more promising for the proponent of the self-defense solution. But again, it proves to make things even worse. For suppose that we first establish the level of threatened harm that would best shift harm away from law-abiding citizens in the case of unprovoked attacker. Then, once again, whatever level of harm this is, it will be necessary to threaten an even greater level of harm to shift the same amount of harm away from law-abiding citizens in a provoked attack, and for the same reason: when people are provoked, they are prone to act more short-sightedly than when they are not. As a result, on this approach, the self-defense solution will justify imposing greater punishment on provoked than on unprovoked offenders. And so, on both the autoretaliator argument and the distributive justice argument, the self-defense solution has a problem: either it treats provoked and unprovoked offenses the same way, and so must justify imposing the same punishment, or it treats them separately and so must justify imposing greater punishment for provoked offenses than for unprovoked offenses. In neither case can the self-defense solution accommodate the belief that offenders who are provoked merit less punishment rather than more. And so, in either case, the self-defense solution seems incapable of passing a reasonable application of the foundational test.

4.4.5 The Harm versus Punishment Objection

I have argued that the self-defense solution fails to draw the line between those we may harm and those we may not in the way required by punishment, and that it justifies harming some offenders too much and others too little. But let us now suppose that I have been mistaken about

all of this and that the solution allows the state to harm all and only offenders, and in proportion to the severity of their offenses, even when mitigating circumstances are taken into account. Even if all of this is so, there is still one final reason to reject the self-defense solution: the solution would justify doing some acts that harm people who break the law, but it would not justify doing acts in order to harm them. Punishment, as we have seen at a number of points, involves not merely harm but intentional harm. And so, like several of the other solutions considered, the self-defense solution cannot justify punishment, as opposed to simply justifying harm, and would therefore fail the entailment test.

To see how this problem arises for the self-defense solution, consider cases in which the principles appealed to by proponents of the self-defense position would justify acts that harm enemy soldiers in national self-defense. Quinn's autoretaliator argument, for example, maintains that it is permissible to install and activate mechanical devices that harm those whose actions trigger them, provided that one installs and activates the devices in an attempt to deter others from doing the wrongful acts that would trigger the devices. This argument could therefore be used to justify one nation's planting land mines at its border to deter another nation from invading it. Montague's distributive justice argument maintains that when a first group's actions make it inevitable that either it or a second group will suffer harm, it is permissible to distribute the harm to the first group. This argument could therefore be used to justify one nation's use of a missile defense system that would destroy an enemy missile shortly after it was launched, even if the result was that harmful debris rained down on the aggressor nation's soldiers.

Now in both cases, it should be clear that the resulting harm to the enemy soldiers is foreseen but not intended. Indeed, Quinn, at least, is explicit about this in the case of the mechanical devices he imagines. The harm that is brought about when a wrongdoer triggers one of the devices, he says, is simply the foreseen "consequence" or "byproduct" of the initial act of activating it (1985: 59, 61). The same, for that matter, must be true of the spiked fences he appeals to in generating his defense of the autoretaliator in the first place. The homeowner who constructs such a fence does so not to harm people but to deter them from trespassing, while foreseeing that those who nonetheless trespass will be harmed. Similarly, when a missile defense system is used to deflect harm from an enemy attack, any resulting harm from the act of deflection is merely a by-product of the system. And yet, the self-defense solution to the problem of punishment depends crucially on the claim that activating a system of legal punishment is analogous to activating these defensive systems, what Quinn calls the "functional equivalence thesis." And this, in turn, generates a dilemma from which the proponent of the self-defense solution cannot escape. If the attitude the solution permits the state to take

toward the harm that befalls people who "trigger" a sentence of five years in jail by breaking the law really is the same as the attitude it permits the state to take toward the harm that befalls soldiers who trigger a land mine by crossing its border or the harm that a trespasser triggers by climbing a spiked fence, then although the practice that the solution justifies will foreseeably harm those who break the law, it will not cause intentional harm and so will not count as punishment. If, on the other hand, the state acts with the intent of harming those who break its laws, then its attitude toward those who break the laws will be crucially different from its attitude toward those who foreseeably trigger harm to themselves – the functional equivalence thesis, that is, will prove to be false – and the argument for the self-defense solution will collapse. To return once more to the relatively simpler case of the spiked fence, if we understand the harm caused to the person who climbs the fence to be merely a foreseen consequence of constructing the fence, then there is no way to move from the assumption that building the fence is permissible to the conclusion that punishment is permissible. If, on the other hand, we understand the harm caused to the person who climbs the fence to be intentional (if, for example, the fence is built to harm him so that he will serve as a cautionary lesson for other would-be trespassers), then to assume that it is permissible to construct the fence is to assume that it is permissible to punish the trespasser rather than to provide an argument for the claim that it is permissible. Either way, the argument from self-defense cannot pass the entailment test and thus cannot provide a satisfactory solution to the problem of punishment.

4.5 HYBRID SOLUTIONS

One final possibility is likely to be raised at this point. For even if it is true that none of the solutions I have considered in this book is successful on its own, it might still be true that some set of such solutions will prove acceptable when suitably combined. Rather than attempting to defend punishment on purely consequentialist or purely retributivist grounds, for example, one might attempt to defend the practice by appealing to both consequentialist and retributivist considerations. Indeed, such hybrid solutions to the problem of punishment are common in the literature, and a defender of punishment might complain that in treating each solution to the problem of punishment in isolation from all of the others, I have been unfairly stacking the deck against the belief that punishment is justified. Perhaps the hybrid approach is not merely the last resort to be appealed to when all other solutions to the problem of punishment fail, that is, but is rather the first and strongest line of defense to be raised in the face of the moral problem that punishment poses. While blending together the most attractive elements of different approaches is certainly appealing in theory, however, it proves unsuccessful in practice. I will focus here on the attempt

to produce a satisfactory hybrid of consequentialist and retributivist considerations, since these are the two most prominent approaches in the literature, but I will also attempt to show that the problem with this approach is ultimately a problem for any hybrid approach.

So suppose that one finds at least something appealing about two particular solutions to the problem of punishment but also agrees that neither approach in itself is fully satisfactory.[40] How, precisely, is one to go about combining them? In attempting to answer this question, it may help to think of each of the solutions in question as a theory that picks out a particular property and that then maintains that the act of punishing an offender has this property and that the fact that it has this property is in itself sufficient to render the act of punishment morally permissible.[41] We can put this point generally as follows:

> Theory 1: If the act of punishing *X* has property *a*, then punishing *X* is morally permissible.
>
> Theory 2: If the act of punishing *X* has property *b*, then punishing *X* is morally permissible

where *a* and *b* stand for such properties as increasing social utility, being permitted by a utility-maximizing rule or practice, being fair, being deserved, being consented to, being reprobative, being educational, and so on. The question then becomes: how can we produce a third theory that makes use of both properties?

4.5.1 Conjunctive Hybrids

The most natural answer to this question is that we can produce such a hybrid by appealing to the conjunction of the two properties. We can say,

[40] For purposes of simplicity, I present my arguments against hybrid solutions in terms of hybrids that seek to combine two solutions to the problem of punishment. If my arguments are correct, however, it should be clear that they apply equally to hybrids involving any number of components. A four-part hybrid that seeks to combine theories 1, 2, 3, and 4, for example, can simply be converted into a two-part hybrid that seeks to combine one theory, which is a hybrid of 1 and 2, with a second theory, which is a hybrid of 3 and 4. If my arguments against hybrid theories apply to all two-part hybrids [such as, e.g., those of Hart (1968a) and Strong (1969)], therefore, they will apply to all more complex hybrids as well [see, e.g., Matravers (2000), which attempts to justify punishment by appealing to a combination of consequentialist, retributivist, moral education, and reprobative considerations, or Braithwaite (1989), which appeals to consequentialist, moral education, and reprobative considerations in justifying punishment, as well as such further writers as Doyle (1967), Kidder (1982), Reitan (1996), and Blumoff (2001), who appeal to a similarly complex combination of considerations].

[41] Note that this formulation does not beg the question against, e.g., the rule-utilitarian solution, because we could simply say that each act of punishing *X* conforms to the rule that a rule-utilitarian would select.

that is, that punishing *X* is morally permissible if the act of punishing *X* has both properties *a* and *b*. So, in the case of the act-utilitarian and desert-based retributivist solutions, for example, a defender of punishment might maintain that even though we can't harm a person just because this would be useful or just because he deserves to suffer, we can harm a person if his suffering would be both useful and deserved. There are two ways in which such a conjunctive hybrid might be developed. They should be understood in different ways and rejected for different reasons.

One conjunctive hybrid arises in the following manner: we believe that property *a* is sufficient to justify punishing someone and that property *b* is sufficient to justify punishing someone, but we recognize that not every act of punishing an offender has property *a* and not every act of punishing an offender has property *b*. In short, we attempt to construct a hybrid of two solutions, each of which is, when taken individually, subject to the not punishing the guilty objection. If the arguments presented in the previous chapters are correct, for example, then this is what we would be doing if we combined utilitarian and desert-based considerations: some offenders do not deserve to suffer, and some punishments of offenders are not (or are not maximally) useful. If we construct a conjunctive hybrid in this manner, however, the resulting solution will be unacceptable regardless of the content of the two separate solutions. Every act of punishing an offender that lacks property *a* will necessarily lack the conjunctive property *a* and *b*, and every act of punishing an offender that lacks property *b* will necessarily lack the conjunctive property *a* and *b*. In every case in which punishment is not useful, for example, it will also be true that punishment is not both useful and deserved. And so, any conjunctive hybrid will fail if at least one of its conjuncts is subject to the not punishing the guilty objection. The resulting hybrid will be subject to the objection as well, so it cannot pass the entailment test that any satisfactory solution to the problem of punishment must pass.

The only possible alternative is to construct a conjunctive hybrid out of two solutions, each of which is immune to the not punishing the guilty objection. The defender of the hybrid approach, that is, must seek to conjoin two theories, each of which points to a property that is present in every single case of punishing an offender. Every act of punishing an offender must have property *a*, that is, and every act of punishing an offender must have property *b*. On this version, of course, neither property is considered sufficient to justify punishment. For if either property is sufficient, then there will be no need for a hybrid in the first place. So, in this scenario, the hybrid theorist must claim that every act of punishing an offender has property *a*, every act of punishing an offender has property *b*, and that while neither *a* nor *b* suffices to justify punishing someone, the conjunction of the two properties does suffice.

There are two reasons to reject any such solution. The first is that, if my objections to the previous solutions have been successful, there are no such theories available to be combined in this manner in the first place. This is so because there are no solutions that successfully entail, even on their own terms, that every offender should be punished. This second kind of conjunctive hybrid, that is, can succeed only if it combines a pair of solutions each of which, taken individually, passes the entailment test, and I have argued that there is no solution that, taken individually, does this. For any version of consequentialism, for example, I have argued that there will always be at least some cases in which punishing an offender would not produce the best consequences. Some offenders do not deserve to suffer and some are not free riders. None have forfeited their rights and none have consented to be punished. And even if all offenders may be resisted in self-defense, this justifies harming them but not punishing them. Since there is no pair of theories each of which appeals to a property of every act of punishing an offender, there are no two theories that can be used to form a conjunctive hybrid in this second version.

But suppose that there were. There is a second problem with any attempt to form a conjunctive hybrid along these lines. The problem arises because there seem to be two different kinds of properties available for any particular solution to the problem of punishment to select. One kind of property identifies a respect in which punishing an offender would make the world a better place from a morally relevant point of view. All things considered, for example, the world is a better place, morally speaking, when it contains more happiness than less, when benefits and burdens are distributed more fairly rather than less, when people generally get what they deserve than when they do not, and so on. The other kind of property that a solution to the problem of punishment might select is one that identifies a way in which a person can lose a kind of moral protection that he formerly had. The properties of forfeiting a right and consenting to give up a right are examples of this second kind of property.

Now suppose that a defender of punishment seeks to produce a conjunctive hybrid from two solutions, each of which appeals to a property of the first sort. In that case, it is utterly mysterious how the conjunction of the two properties could justify punishment if the conjuncts taken individually do not. If the fact that harming you would increase the happiness in the world does not make harming you permissible, for example, and if the fact that harming you would make the world contain more fairness does not make harming you permissible, then how could it be permissible to harm you merely because this would create both more happiness and more fairness? Either it is permissible to harm you as a means of producing a morally better outcome or it is not. If it is, then there is no need for the hybrid in the first place. And if it is not, then the conjunctive hybrid will be no more successful than its individual conjuncts.

But now, suppose that a hybrid theorist instead seeks to use at least one solution that appeals to a property of the second sort. Perhaps, for example, she seeks to combine forfeiture-based retributivism with act-utilitarianism or the consent solution with considerations about people getting what they deserve. If this is the case, then it is again difficult to see how the conjunction of the two solutions can succeed if the conjuncts fail when taken individually, though for a different reason. For in this case, either we think that the offender's act has already stripped him of his rights or we think it has not. If we think it has, then it is again difficult to see the need for a hybrid theory. If we think it has not, then it is difficult to see how combining the theory with a second theory could change our minds. If your act of stealing a television set does not in itself mean that you have forfeited five years of freedom, for example, then how could the added fact that putting you in prison for five years would be useful (or would educate you, or would give you what you deserve) mean that you have forfeited this right? While combining two solutions to the problem of punishment in this way may seem attractive in theory, therefore, it proves unacceptable in practice.

4.5.2 Disjunctive Hybrids

A defender of punishment who is attracted to the hybrid approach could agree that conjunctive hybrids are unsatisfactory but attempt to combine two solutions in some other manner. The obvious alternative is to produce a disjunctive hybrid, on which it is permissible to punish an offender if the punishment would have either property *a* or property *b*. A defender of punishment might argue, for example, that punishing a given person is justified so long as punishment is either deserved or useful. This approach will fail for two reasons. First, there can be cases in which punishing an offender might be neither useful nor deserved. So, there will be cases in which the defender of punishment is committed to saying that it is permissible to punish the offender but the disjunctive hybrid fails to support this claim. The hybrid, in short, will again be subject to the not punishing the guilty objection and so will again fail the entailment test. Second, and more importantly, the disjunctive approach to constructing a hybrid will generate the punishing the innocent objection. If punishment is justified so long as it is either useful or deserved, for example, then it will be justified in many cases in which it is useful (or deserved) even though the person has committed no legal offense in the first place. A disjunctive hybrid could avoid the punishing the innocent objection only if both of its disjuncts were theories that, taken individually, avoided the objection. But I have argued in the past three chapters that there are no theories that, taken individually, avoid the punishing the innocent objection. If this is so, then there is no disjunctive hybrid that avoids the

objection either. And if this is so, then there is no disjunctive hybrid that can pass the entailment test, let alone the foundational test.

Neither disjunctive nor conjunctive hybrids provide acceptable solutions to the problem of punishment. Is there some further possibility? So far as I can see, there is not. So far as I can see, therefore, there is no solution to the problem of punishment at all. Virtually everyone believes that the state's practice of punishing people for breaking the law is morally permissible. But it turns out that we have no good reason to believe that this is true.

5

The Appeal to Necessity

5.0 OVERVIEW

The problem of punishment arises from the fact that when the state punishes people for breaking the law, it subjects them to various treatments that are intentionally harmful. Typically, it is impermissible to do an act with the intention of harming another person. A successful solution to the problem of punishment must therefore explain why the fact that a person has broken a just and reasonable law makes it morally permissible for the state to treat him in ways that would otherwise be impermissible. In Chapter 1, I argued that this problem posed by the practice of punishment is a genuine moral problem, and in Chapters 2–4, I argued that none of the attempts to solve the problem that have been offered in the voluminous literature on the subject are successful.

Let us now suppose that I have been correct about all of this. What should we do? At a general level, there are only two possible answers to this question: we can continue to punish and hope that a satisfactory defense of its permissibility will eventually emerge, or we can abolish the practice, at least until a satisfactory defense of its permissibility can be provided.[1] The first option strikes me as morally repugnant and the second as morally imperative. In no other realm of human interaction would we allow one group of people to intentionally inflict serious harm on another if no satisfactory justification for the moral permissibility of this practice was available. So, the answer to the question of what we should do if we can find no solution to the problem of punishment would seem to be simple: we should stop punishing people for breaking the law.

[1] Though see Teichman for an attempt to carve out a third possibility on which "the institution of punishment cannot be shown to be unjustified, and cannot be shown to be justified either" (1973: 336).

But many people believe that things are not so simple. Punishment, after all, is the cornerstone of our justice system's response to violations of the law. It is difficult to imagine a system of justice without it. Thus, even those who are skeptical of the various solutions to the problem of punishment typically balk at the seemingly straightforward implication that we should abolish it. As one such critic of these solutions has put it, "we cannot break away from punishment. . . . as long as rules exist, so will punishment" [Bean (1981: 193)]. This worry that we simply cannot do without punishment gives rise to one final position that must be confronted. Although this position does not purport to provide a straightforward defense of the claim that punishment is morally permissible, it does purport to block the move from the absence of such a defense to the conclusion that punishment should be abolished. It does this by appealing to the claim that punishment is necessary.

The appeal to necessity that I have in mind here is not the claim made by the logical entailment argument that punishment is logically necessary (see Section 1.2.3), the claim that the permissibility of punishment is entailed analytically by the mere acceptance of legal rules. Rather, the appeal amounts to the claim that punishment is practically necessary: that given the way the world actually is, punishment is needed for society to exist at all or, at least, as Hill puts it in attributing this view to Kant, that it is "necessary to provide even the minimum conditions for just mutual relations" [(1997: 195); see also Hill (1992: 32)]. This appeal to necessity need not be understood as providing a full-fledged justification for punishment. Indeed, I am inclined to think that even if the appeal is successful, it would at best provide an excuse for engaging in the practice despite its impermissibility rather than a reason for concluding that it is morally permissible. If one community could demonstrate that it was necessary for it to conquer and enslave another community, after all, either for survival or for providing the minimal conditions needed for its members to enjoy just mutual relations among themselves, for example, I suspect that we would not think this would give the community the right to do so, though we might think it would render their immoral behavior excusable. But whether we consider the appeal to necessity to represent one final solution to the problem of punishment or an excuse for us to continue engaging in it even if we cannot find a solution, it raises the same basic question: is punishment necessary? And so, on either version, the only way to overcome the appeal to necessity is to identify and defend at least one acceptable of doing without it. This is the goal of this final chapter.

One way we could do without punishment would be to replace it with something else. Some people, for example, have argued that punishment should be replaced by treatment and therapy.[2] A consistent believer in

[2] See, e.g., Skinner (1953), Baylis (1968a: 46–8), Menniger (1968), and Blume and Blume (1989), criticisms of this view in general by McCloskey (1978) and of Baylis in particular by Koehl (1968) and Regan (1968), as well as a reply by Baylis (1968b).

just deserts might argue that legal punishment should be replaced with a system in which bad people are made to suffer (and good people made to flourish) whether or not they break the law. A pacifist might argue for an entirely new kind of legal system that would "require a whole new culture, a whole new system of socialisation, a whole new system of values and a whole new political organisation within which the system could be put to work" [Fatic (1995: 202)][3] and so on. I am skeptical about all of these suggestions. But since none of them are inconsistent with the response to the appeal to necessity that I wish to advance here, I will not argue against them.

The second way we could do without punishment involves not replacing it with something new, but rather relying more heavily than we currently do on something old: victim restitution.[4] Nothing that has been said in this book against punishment, it is important to emphasize, counts in any way against compulsory victim restitution. The problem of punishment arose for two reasons: punishment involves the state's intentionally harming some of its citizens, and it involves treating the line between those who break the law and those who do not as justifying treating people on one side in ways that it would not treat those on the other. But neither of these features of punishment is a feature of compulsory victim restitution. As we saw in Section 1.1.8, although compulsory victim restitution typically does involve predictable harm to the offender, pure restitution does not involve harming the offender intentionally, either as an end in itself or as a means to a further end. And while punishment involves the state's treating people who break the law in ways that we would not permit it to treat people who do not break the law, compulsory victim restitution does not. The state compels non-offenders to make restitution to others all the time. Whenever one party successfully sues another party for damages in a civil lawsuit, for example, someone who is guilty of no violation of the law is nonetheless compelled to compensate someone else for damages she is found to have caused. Compelling an offender to make restitution to his victims, therefore, does not raise the difficult moral problem that punishment does. And the claim that the state has no right to punish people for breaking the law therefore provides no reason to suspect that the state has no right to

[3] See also Fatic (1995: 203, 222, 226). Luke (1996) defends a similar view on which punishment could be abolished given sufficiently powerful changes in the educational and social environments.

[4] As Strang points out, "Traditionally, in most places throughout the world ... compensation and restitution have been the dominant model of conflict resolution" (2002: 3), and it is only in the modern state that societies have come to rely more and more on punishment and less and less on victim restitution (3–5, 193). A similar point is noted by Strickland (2004: 2). See also Abel and Marsh (1984: 25–9) and Schafer (1970: chaps. 1, 2).

compel offenders to make restitution to their victims.[5] The second way to do without punishment, then, is to stop punishing people for breaking the law and do more to compel offenders to make restitution to their victims. When a person breaks the law, on this approach, the state should force him to compensate his victims for the harms that he has wrongfully caused them, but it should do nothing more.[6]

This second way of doing without punishment must not be confused with the claim that victim restitution should play a more important role in the justice system. The claim that we should pay more attention to victim restitution than we currently do is relatively modest. The claim that by doing so we could do without legal punishment altogether is not. And it is the latter claim that I wish to examine in this concluding chapter as a means of responding to the appeal to necessity. The claim that we should do without punishment by relying on a system of compulsory victim restitution was first proposed briefly in the modern literature on punishment by del Vecchio (1965), but it is most widely associated with the article "Restitution: A New Paradigm of Criminal Justice" (1977) by Randy Barnett.[7] Following Barnett, I will refer to this view as the "theory of pure restitution" (1978: 220).[8] I will begin, in Sections 5.1 and 5.2, by

[5] This is not to say that there are no problems in accounting for the state's right to harm people by compelling them to make restitution to their victims. Surely there are problems [see, e.g., Montague (2002: 3, 5–6)], just as there are problems in accounting for the state's right to harm people by levying taxes, compelling them to serve on juries, requiring them to obey zoning regulations, and so on. But these problems can be passed over here, since whether or not they can be solved has no bearing on the central thesis of this book: that it is morally impermissible for the state to punish people for breaking the law.

[6] This is not to insist that the state should never respond to wrongdoing by providing treatment for the offender. As I will argue in Section 5.3.3, it may be possible under certain circumstances to justify mandatory treatment of an offender on restitution grounds.

[7] Barnett (1977). Barnett also defends this position in (1980; 1985: 63–7; 1998). Some writers have attempted to defend Barnett's proposal from some specific objections that have been raised against it [e.g., Hajdin (1987), Wilkinson (1996), Long (1999: 125–40), Ellin (2000)]. Wilson, although in a more limited context (1983: 521–3), and others have argued that restitution should play a much greater role in the criminal justice system than it currently does [e.g., Wright (1996)], but very few people other than Barnett have defended the claim that restitution should replace punishment entirely [Abel and Marsh come very close to doing so, but at several points it becomes clear that their proposal still permits punishment in some cases, e.g., to deter in some instances and to punish offenders who refuse to provide restitution in others (e.g., 1984: 171, 178, 184)].

[8] For the distinction between pure restitution and punitive restitution, see Section 1.1.8. The theory of pure restitution is closely related to what is now commonly referred to as the "restorative justice" movement. Proponents of restorative justice, like defenders of the theory of pure restitution, maintain that the legal system should focus primarily on repairing the various harms that are done when laws are broken. The term restorative justice, however, is generally taken to refer to a wider set of views than this, including views about the effectiveness of mediation and of meetings between victims and offenders

clarifying both the limited sense in which I mean to defend the theory of pure restitution and the content of the theory itself. Then, in the sections that follow, I will attempt to provide a defense of the theory, so understood, from the wide variety of serious and potentially decisive objections that have been raised against it by such writers as Miller (1978), Pilon (1978), Kleinberg (1980), Dagger (1991), and Tunick (1992: 156–62). The few writers who have sought to defend the theory of pure restitution have attempted to respond to at least some of these objections, and have done so with varying degrees of success. But no one has taken on all the objections raised against the theory and established convincingly that they can be successfully overcome.[9] If all of the objections to the theory of pure restitution can be overcome, of course, this alone will not suffice to demonstrate that a system of compulsory victim restitution is the best alternative to a system of punishment. There are many possible alternatives to punishment, and some of them may turn out to be even better. But overcoming the objections to the theory will establish that we can do without punishment, and establishing just this much will be enough for the purposes of this chapter. If we can do without punishment, after all, then the appeal to necessity must be rejected. And if the appeal to necessity must be rejected, then our inability to provide a satisfactory solution to the problem of punishment means that we should end the practice of punishing people for breaking the law. I have argued, in short, that we have no good reason to believe that punishment is morally permissible. If, in addition, we have no good reason to believe that

[see, e.g., Moore (1993), Sullivan and Tifft (2001: chap. 3), Strickland (2004: 1–2)] and about the nature and causes of lawlessness [see, e.g., Sullivan and Tifft (2001: chap. 7)], subjects about which the theory of pure restitution is neutral. In addition, there is considerable debate within the restorative justice movement about whether or not it should exclude punishment entirely [see, e.g., Strang (2002: 203–4), Zehr (2002: 12–13, 58–9), and the various essays collected in von Hirsch, Roberts, and Bottoms (2003)], whereas the theory of pure restitution rejects punishment altogether. While the thesis of this chapter can therefore be taken to support one key value of the restorative justice movement, I will continue to state the thesis in terms of a defense of the theory of pure restitution rather than in terms of restorative justice [for defenses of two closely related additional views, though ones that explicitly reject the restitution approach as an alternative to punishment, see also Fatic (1995) and Sayre-McCord (2001, 2002)].

9 In saying this, I mean to include Barnett himself, whose responses to some important objections are criticized later, as well as Abel and Marsh (1984), who devote surprisingly little space to a number of important objections and whose responses to some of the objections they do address are criticized at several points later. The same is true of the more recent book by Golash, which argues in some detail against several justifications of punishment but offers only a brief defense of restitution as one small part of a nonpunitive alternative system of justice (2005: 162–6). Ellin's article (2000) attempts to fend off a larger number of objections to the theory than do most of its defenders, but at just under seventeen pages in length, it does not go into sufficient depth when discussing many of them.

punishment is practically necessary, then we have no good reason to resist the conclusion that we should simply abolish punishment.

5.1 THE THEORY OF PURE RESTITUTION AS A RESPONSE TO THE APPEAL TO NECESSITY

The theory of pure restitution consists in the conjunction of two claims: the claim that the state should not punish people for breaking the law and the claim that the state should compel people who break the law to compensate their victims for the harms they have wrongfully caused them. The theory can therefore be attacked from two directions. From one side, the theory can be attacked by those who believe that it is impermissible for the state to compel people to compensate those they wrongfully harm. This belief can be aimed at the theory in two ways. First, a critic could hold that compulsory victim restitution is impermissible but maintain that punishment is permissible. Whereas the theory promotes restitution without punishment, this critic of the theory would endorse punishment without restitution. It is extremely difficult, however, to imagine any argument for this position, and I am not aware of anyone who has held it. I will therefore ignore it. But a critic who attacks the theory from this first direction could also use the belief that compulsory victim restitution is impermissible in a second way. On this version, the critic would oppose both restitution and punishment. It is easier to understand how there may be credible versions of this critical position. One might hold, for example, that there should be no state at all, or that while it is permissible for the state to attempt to prevent wrongful harms from occurring, it should do nothing once such harms occur, or aim entirely at rehabilitation, or aim at forgiveness, and so on. But while much may be said in defense of some of these positions, this need not occupy us here. The thesis of this book is that punishing people for breaking the law is morally impermissible, and I am concerned with defending the theory of pure restitution here only insofar as it serves as a response to what I am calling the "appeal to necessity," the claim that we may permissibly (or, at least, excusably) continue to punish because it is practically necessary. Since this second version of the first kind of critic of the theory agrees that we can do without punishment, he rejects the appeal to necessity. This kind of critic thus poses no threat to the thesis that this book seeks to defend. If it turns out that this critic of the theory of pure restitution is correct, punishment will still prove to be morally impermissible. I will therefore ignore this critic as well. I will simply assume, as most people will surely concede in any event, that it is morally permissible for the state to compel offenders to compensate their victims. I will also assume, again without argument, that it is permissible for the state to compel defendants in civil cases to make restitution when courts rule in favor of plaintiffs. If a

jury determines that a doctor has negligently harmed her patient, for example, then it is permissible for the state to compel her to pay compensatory damages. These assumptions are reasonable given the limited respect in which I want to defend the theory of pure restitution. If the assumptions turn out to be mistaken, then the theory will have to be rejected, but so will the practice of punishment. And so, for the purposes of this book, the first kind of critic of the theory can be passed over entirely.

There is, however, a second kind of critic. This critic agrees that compulsory victim restitution is morally permissible but believes that punishment is as well. If the position of this second critic can be sustained, then the thesis of this book must be rejected. And so, this second critic must be confronted head on. This second critic can attempt to justify his position either directly or indirectly. He can attempt to offer a positive defense of punishment, that is, or he can attempt to identify unacceptable implications of the claim that restitution should be practiced without punishment. I have already implicitly attempted to defend the theory from the first attack in the first four chapters of this book. In Chapter 1, I argued that the problem of justifying the moral permissibility of punishment required a solution, and in Chapters 2–4, I argued that we have no such solution. In the remainder of this chapter, I will therefore focus entirely on the second attack. A number of people have argued that the theory of pure restitution should be rejected on the grounds that restitution without punishment has unacceptable implications. I will attempt to show that these objections can all be overcome.

Before responding to the various *reductio ad absurdum* objections that have been raised against the theory of pure restitution, however, it is important to distinguish between two ways in which such objections might be offered. First, a critic might claim that the theory of pure restitution has a particular unacceptable implication and might maintain that punishment does not have this implication. Second, a critic might claim that the theory of pure restitution has a particular unacceptable implication and admit that punishment does as well. The former critic rejects the theory because he believes in punishment along with restitution, while the latter rejects it because he believes in neither punishment nor restitution. As I have already noted, I will not be concerned in this chapter with the latter critic. The thesis of this book is that it is morally impermissible for the state to punish people for breaking the law, and the latter critic agrees with this thesis. I will thus focus exclusively on the former critic. This is the only critic whose attack on the theory of pure restitution poses a challenge to the thesis of this book. And because of this feature of the dialectic of the argument, there are not two but three ways that a defender of the theory of pure restitution in this limited context can attempt to respond to the *reductio ad absurdum* objections that will be considered

here: he can attempt to show that the theory does not have the implication in question; he can attempt to show that the implication is not unacceptable; or he can attempt to show that if the theory does have the implication, then so does the practice of punishment. My claim in this chapter, then, is that there is no objection to the theory of pure restitution that cannot be overcome in at least one of these three ways. If this is so, then there is at least one way that we can manage acceptably without resort to punishment. And if this is so, then the appeal to necessity must be rejected and, along with it, the practice of punishment.

5.2 CLARIFYING THE THEORY OF PURE RESTITUTION

The theory of pure restitution can be put as follows: when offenders break the law, they cause wrongful harms to their victims. When people cause wrongful harms to their victims, this generates a debt: they owe their victims compensation sufficient to restore them to the level of well-being that they rightfully enjoyed prior to being wrongfully harmed. People who commit such offenses therefore owe such compensation to their victims. When people break the law, it is thus morally appropriate for the state to compel them to make such compensation. When an offender is forced to compensate her victims for the harms that she has wrongfully caused them, the offender is harmed as a result in various ways. But the state does not extract compensation from the offender in order to harm her, either as an end in itself or as a means to some further end. It extracts compensation to compel her to restore what she has wrongfully damaged with the understanding that this will (typically) result in harm to the offender. The fact that an offender has broken the law, that is, does justify the state's acting in ways that will predictably harm her, on this account, but it does not justify acting in ways that aim at harming her. Punishment aims at harming the offender, while compulsory victim restitution does not. And so, according to the theory of pure restitution, it is morally permissible for the state to extract victim compensation from people for breaking the law, even though it is not morally permissible for the state to punish them.[10]

A few clarifications are required before we consider the various objections raised against the theory. First, the theory is restricted to cases in

[10] For a more complete justification of the claim that restitution is importantly different from punishment in this respect, see Section 1.1, especially Sections 1.1.4 and 1.1.8. In following Barnett's terminology, here, it may be useful to note, the word "restitution" is being used in the common sense of restoring the victim to his previous level of well-being. The theory should therefore not be confused with the more narrow and technical sense of restitution used in tort law involving cases in which a defendant unjustly benefits at the expense of a plaintiff, who is then entitled to sue for restitution of the benefit [see Kionka (1999: 385)].

which offenders *harm* their victims. For the purposes of this chapter, I will say that an offender's act harms a victim if the act makes the victim worse off than he would have been had the act not occurred. The consequences of some acts are uncontroversially harms in this sense. If my act causes you physical pain, for example, or damage to your property, or loss of money, then my act clearly harms you. The status of other sorts of consequences is less clear. If my act merely offends you, for example, or upsets you, or depresses you, should this count as my harming you or simply as your happening to respond to my (nonharmful) act in certain ways that are unpleasant for you? Since the theory of pure restitution applies only to cases in which an offender has caused harm, it might at first seem that a defender of the theory would owe us a detailed answer to such questions. But this is not so. The question of which cases should count as harms can simply be set aside. If the most reasonable account of what it is to harm someone ends up including, for example, acts that merely cause offense, then the theory will say that such acts potentially ground a claim of restitution. If the most reasonable account of what it is to harm someone ends up excluding such acts, then the theory will say that such acts cannot be used as a basis for restitution. There can therefore be different understandings of what the theory implies in particular cases, but this fact does not count against the theory itself. There are, and have been, different understandings of what counts as harm in tort law, after all, but this fact does not itself count against the practice of torts.[11] The theory, in any event, is not designed to answer the question "which people have caused wrongful harm?" but rather "what should be done when people have caused wrongful harm?"

Second, the theory is restricted to cases in which offenders *wrongfully* cause harm. For the purposes of this chapter, I will say that an offender's act wrongfully harms a victim if the offender's harmful act is prohibited by law[12] and if the legal prohibition is just and reasonable.[13] As with the

[11] For some interesting examples of questions involving harm in tort law, see Kionka (1999: 150, 156, 349, 350, 424, 454). For an illuminating discussion of the ways in which the answers to some of these questions have evolved, see White [(2003: e.g., on changing views about whether or not to include purely mental harms: 102–4)].

[12] By being restricted to cases in which an offender violates the law, the theory should not be taken to imply that the state may never extract restitution when someone does not violate the law. The theory does not, for example, deny that it is permissible for the state to compel a doctor to pay restitution to a patient in a civil malpractice suit. The stipulation here simply indicates that the theory of restitution is presented as an answer to the question of what the state may permissibly do to those who violate the law.

[13] It should be noted that an act could prove to be "wrongful" in the sense used here without being morally wrong. It might, for example, be morally permissible for me to break into your cabin and burn some of your furniture if this is necessary for me to avoid freezing to death in the wilderness, while the laws that protect your property from such invasion and destruction might at the same time be just and reasonable. In such a case,

question of the nature of harm, a defender of the theory of pure restitution need not provide a complete answer to the question of which acts should be considered wrongful in this sense. The question the theory seeks to answer is not "what acts should be legally prohibited?" but rather "what is the state entitled to do when people perform acts that are (justifiably) legally prohibited?" In examining some of the implications of the theory, I will assume, at least for purposes of illustration, that acts that are uncontroversially illegal (rape, murder, arson, theft, fraud, etc.) are wrongful in this sense and that acts that are uncontroversially legal (putting your left shoe on before your right, reading the newspaper, etc.) are not wrongful in this sense, but nothing of substance will turn on the truth of these assumptions.

Suppose, for example, that you own a small, marginally competitive independent bookstore. If I open a larger, more competitive branch of a chain bookstore next door, this may drive you out of business and thus harm you. But on the assumption that it is legal for me to open such a bookstore, the harm I cause you is not wrongfully caused in the sense in which I am using that term, and so, according to the theory of pure restitution, I would owe you no restitution. If, on the other hand, I burn down your bookstore, that may also drive you out of business and harm you to a comparable degree. On the assumption that it is justifiably illegal for me to burn down your bookstore, this harm would be wrongfully caused and I would therefore owe you restitution. The assumption that the law as it stands is, in general, morally justified is useful in illustrating the theory of pure restitution. But it is not necessary to defend the theory. If a solid case could be made for a law that would prohibit my opening the bookstore next door to you (say, on antimonopoly or zoning grounds), for example, then the theory will say that I would owe you restitution for my actions. If a successful case could be made against laws prohibiting arson, then under the theory I would owe you no restitution for burning down your bookstore. But if these implications were considered objectionable, this fact could not be taken as evidence against the theory of pure restitution. They could only be taken as reasons to doubt the merits of the laws in question.

Third, the theory says that offenders must make restitution to their victims when they are *responsible* for the harms they have wrongfully caused. As with the question of which consequences should count as harms and which harms should count as wrongful harms, some judgments about responsibility will be relatively uncontroversial, while others will be

the theory of pure restitution accommodates both the intuition that morality permits me to act as I do and the intuition that you are entitled to compensation for the harm done to your property [for illuminating discussions of this case, see, e.g., Feinberg (1978), Thomson (1980), and Montague (1984)].

much more difficult. Assume, for example, that it is and should be illegal for me to punch you in the nose and that punching you in the nose uncontroversially harms you. If I am of sound mind, spend weeks planning to catch you so that I can punch you in the nose, execute this plan flawlessly, and succeed in punching you in the nose, it should seem clear that I am responsible for having wrongfully harmed you. If, on the other hand, a mad scientist kidnaps and brainwashes me into punching you in the nose, then it should seem clear that although you have been wrongfully harmed, I am not responsible. In the former case, the theory would say that I owe you compensation, while in the latter case it would not. But what about cases in which I punch you in the nose only because I have been under duress or I act in response to extreme provocation? Here there is room for a variety of views about how responsible, if at all, I am for the harm I have caused. And there are many other difficult questions about how much of a resulting harm, if any, an offender can justly and reasonably be held responsible for. Should it matter if a second harmful cause intervenes after the offender's wrongful act and before the victim is harmed? Should it matter if the victim failed to take reasonable precautions to avoid the harm, either before or after the offender's act? Here, once again, the defender of the theory of pure restitution need not take a particular position on which of these views is most reasonable. The theory maintains that an offender owes restitution to the extent that he is responsible for causing wrongful harms, but it can leave open the question of what responsibility means in particular cases. The question that the theory is concerned to answer is not "which people are responsible, and to what degree, for having wrongfully caused harms?" but rather "what is the state entitled to do to those offenders who are found to be responsible (wholly or partially) for having wrongfully caused harms?" And if a critic is tempted to complain that this response is evasive, it should be remembered that I am assuming that compelling restitution in tort law is morally permissible, and tort law does not provide simple answers to these questions either. Indeed, the history of tort law is largely an ongoing debate about how best to answer questions such as this.[14]

Fourth, the theory says that when the offender is responsible for his act, when the act is wrongful, and when the act harms the victim, the state should compel the offender to *restore* his victim to the level of well-being he rightfully enjoyed prior to the offense. What, precisely, would this involve? Ideally, restitution would involve restoring the victim to the condition she enjoyed prior to the offense. To the extent that this is impossible in any particular circumstance, the theory maintains that the offender must restore the victim to a condition equivalent in value to the condition she enjoyed prior to the offense. And, to the extent that this,

[14] For a useful overview, see Kionka (1999).

too, is impossible in any particular circumstance, the theory maintains that the offender must restore the victim to a condition that is as close in value to her original condition as possible. How this is to be done is, once more, a question that the theory need not answer. Just as a theory of punishment need not tell us how, specifically, we should inflict on an offender the suffering he deserves, the theory of pure restitution need not tell us how, specifically, we should have the offender make restitution to his victim.

Finally, the theory says that when these conditions obtain, the offender must restore his victim to the level of well-being that the victim *rightfully* enjoyed prior to the offense. By saying that the victim "rightfully" enjoyed some particular good, I mean that the victim's enjoyment of that good was protected by a just and reasonable law. Suppose, for example, that I break into your house and destroy your computer and television set. If both of these objects belong to you, then the theory dictates that I owe you compensation for both losses. If, on the other hand, you owned the television set but stole the computer from someone else, then the theory would say that I owe you compensation for the television set but not for the computer.

We can therefore summarize the theory of pure restitution as follows: if an offender is responsible for having wrongfully harmed a victim, then (a) the state should compel the offender to restore the victim to the level of well-being that the victim rightfully enjoyed prior to the offense and (b) the state should not punish the offender. If there is no good reason to reject this theory, then there is no good reason to accept the appeal to necessity. And if there is no good reason to accept the appeal to necessity, then there is no good reason to reject the claim that we should abolish the practice of punishing people for breaking the law. I will argue in the remainder of this chapter that there is no good reason to reject this theory that would not also count as a reason to reject the practice of punishment. Whether one ultimately accepts or rejects the theory of pure restitution, therefore, one must reject the practice of punishment.

5.3 THE HARM TO SOCIETY OBJECTION

Perhaps the most common and fundamental objection to the theory of pure restitution arises from its seemingly individualistic nature. "The idea of restitution is actually quite simple," Barnett writes in defending the theory. "It views crime as an offense by one individual against the rights of another" (1977: 219). "Where we once saw an offense against society," he adds, "we now see an offense against an individual victim. In a way, it is a common sense view of crime. *The armed robber did not rob society; he robbed the victim*" (219, emphasis in the original).

5.3.1 The Objection

This may be a commonsense view of what is involved when a person violates the law. But, according to its critics, it is also a deeply mistaken view. An offense committed against a single individual does more than wrongfully harm that individual, the objection maintains; it also has important repercussions for the rest of society. If a gunman robs my neighbor as he is about to enter his home, for example, then the gunman has wrongfully harmed my neighbor. But by robbing my neighbor, he may also wrongfully cause various harms to me. I may suffer from anxiety and lost sleep as a result. I may feel forced to incur the added expense of installing and maintaining a security system in my house or of buying a gun. The value of my property may go down. My insurance rates may go up. I may incur various opportunity costs, forfeiting whatever I would have enjoyed with the time and money that I have instead been forced to devote to responding to the offense against my neighbor. And these negative consequences that the armed robbery of my neighbor cause to me may, in turn, impose further costs on still other people, including people who don't even know that my neighbor was robbed (e.g., people who would have benefited from my money had I not bought the security system instead). In all of these ways, and in many others as well, an offense against a single individual can also seem to be an offense against the rest of society. Yet since, according to the objection, the theory of pure restitution requires only that offenders compensate their individual victims, the theory seems incapable of recognizing and addressing this fact. And so, it is said, the theory must be rejected. Requiring restitution may be good enough when dealing with the victim of the offense, but punishment is still necessary when dealing with the rest of society. As one critic of the theory has put it, "the [pure] restitution theory understates the importance and the complexity of the network of relationships that is disrupted by crime, relationships too complex to be repaired through payment of compensation" [Hoekema (1991: 343)].[15]

5.3.2 The Biting the Bullet Response

There are two different ways in which a proponent of the theory of pure restitution might attempt to respond to the harm to society objection. One is to bite the bullet and accept the implication that the theory cannot satisfy anyone other than the offender's immediate victim. This is Barnett's response. He represents the harm to society objection as maintaining that because an offender wrongs society as a whole, it follows that "society,

[15] See also Hoekema (1991: 340). This harm to society objection has been pressed by a number of other critics of the theory of pure restitution, including Miller (1978: 359), Kleinberg (1980: 277), and Tunick (1992: 158ff.).

that is, individuals other than the victim, deserves some satisfaction from the offender. Restitution, it is argued, will not satisfy the lust for revenge felt by the victim or the 'community's sense of justice.' " He then responds to the objection as follows: "This criticism appears to be overdrawn. Today most members of the community are mere spectators of the criminal justice system, and this is largely true even of the victim" (1977: 225). Barnett therefore agrees that the theory of restitution can do nothing to satisfy the rest of society, but he insists that this is not a problem for the theory since it is true of punishment as well.

This response to the harm to society objection is unsatisfactory. It conflates two importantly distinct claims made by the critic of the theory of pure restitution: the general claim that individuals other than the victim deserve *some* kind of satisfaction from the offender and the more specific claim that what they deserve is, in particular, the satisfaction of their desire for *revenge*. Barnett's reply addresses only the second claim, but it is the first that represents the stronger version of the objection.[16] The claim that an armed robber might incur a debt of some sort to people other than his immediate victim seems plausible enough, and if it is true, then a proponent of the theory must find a better way to deal with it than simply ignoring it.

5.3.3 The Secondary Victims Response

Fortunately for the defender of the theory, such a response is available. The response begins by acknowledging that, at least in a good number of cases, individuals other than the offender's immediate victim are also wrongfully harmed by the offense. Rather than ignoring this fact, as Barnett seems willing to do, this response treats these other people as additional victims, albeit secondary rather than direct or primary victims. What I will call the "secondary victims response" to the harm to society objection, then, insists that in such cases the offender must make restitution to these people as well. The secondary victims response permits the rest of society to be considered as a further victim and then to be treated in the same way that any other victim would be treated by the theory.[17]

[16] The concern that the offender must satisfy the desire for revenge aroused in his victims and others is addressed in Section 5.13.

[17] It might be objected that the secondary victims response depends on a certain view of responsibility, which prevents the defender of the theory of pure restitution from maintaining neutrality about the nature of responsibility. On one view, for example, an offender might only be responsible for the harms that he intends to inflict on his immediate victim; on this view, the theory could not justify restitution to secondary victims. While it is true that the secondary victims response cannot be sustained with a certain view of responsibility, however, it is not true that this poses a problem for the theory. If it is assumed that an offender is not responsible for any of the harms he causes

If punishment is not needed because of the offender's effect on his direct or primary victim, then it is not needed because of his effect on the rest of society either.

How could an offender be made to repay his secondary victims? In an ideal world, the state would simply perform all of the relevant calculations, determine precisely how much harm an armed robber who victimized a particular person had wrongfully caused to the rest of his community, and then assign a precise dollar amount to each instance of harm. Perhaps it would determine that the robber caused more secondary harm to the victim's immediate neighbors, with the level of harm gradually dissipating as people were farther and farther removed from the victim. In this case, the offender would owe some money to each nearby neighbor and then some, but less, to others farther away. Or perhaps the state would determine that the harm had been distributed in some other pattern, in which case the requisite monetary payments would be so distributed as well. So, an initial attempt to sustain the secondary victims response would maintain that the offender should be compelled to give just enough money to every secondary victim to compensate each of them for the harm to them wrongfully caused by his offence. This answer follows from the basic idea of restitution and shows how a theory of pure restitution can accommodate the legitimate demand for satisfaction from the offender that secondary victims may often have without resorting to punishment.

This version of the secondary victims response is also, of course, completely impractical. But this does not mean that the response itself is mistaken. After all, it may often be impractical to punish a criminal by subjecting him to precisely the amount of suffering that he is thought to merit, but this does not mean that punishment is mistaken. It simply means, in either case, that in the real world we often cannot do precisely what we should do or are entitled to do. And here, as in other relevantly similar cases, if the law cannot do precisely what it should or may do, this does not mean that it should or may do nothing at all. It simply means that it should do the best it can to approximate what it should or may do. Consider, for example, the widely publicized lawsuits against the major tobacco companies. Assume, for the sake of the example, that these companies wrongfully harmed millions of smokers and should have to pay damages to them. Ideally, the state would identify every person who has been wrongfully harmed by tobacco smoke, either directly or indirectly, determine the precise dollar amount of the harm that each person suffered, and use this as a basis for determining the precise amount of compensation owed to each of the millions of victims involved. Since it

to society, then it is true that under the theory of pure restitution, the offender owes no debt to society, but given the assumption, this result should seem perfectly reasonable.

would be permissible for the state to do this if it could, but also since, as a practical matter, it cannot, the state can attempt to approximate this result in a reasonable manner. It might, for example, take a lump sum payment from the companies and use it to improve the health care system in ways that would be most likely to benefit those who were most harmed by the tobacco companies. Similarly, then, given that it would be permissible for the state to compel the robber to pay every member of the community precisely the compensation he owed them, but given also that this would be practically impossible, the state can attempt to approximate this result in a reasonable manner. It might, for example, compel the robber to pay a lump sum to the city, which would use the money to hire an extra patrol officer, or two, or three, depending on how much police power was necessary to restore the community to its previous level of well-being, and which would deploy the new officers in the manner most likely to bring the greatest benefits to those who had suffered the greatest harms as a result of the robbery.

Three objections to the secondary victims response merit consideration at this point. The first is that in acknowledging the existence of secondary victims, the response departs too much from the principle of restitution. Dagger, for example, argues that to concede that offenders sometimes wrong society as a whole is "to surrender one of the core notions of pure restitution – the idea that criminals do not incur a debt to society" (1991: 32). Dagger supports this charge by citing Barnett's claim that the offender's debt "is not to society; it is to the victim" (30). But this objection confuses the principle of pure restitution with a particular application of the principle. The principle of pure restitution maintains that an offender incurs a debt to those he wrongfully harms. The principle is neutral on the question of whether society as a whole can be harmed by an offender's wrongful action. Barnett's substantive application of the principle of restitution, it is true, involves appealing to the further claim that society as a whole is not wrongfully harmed by an armed robber. If Barnett is right, then the harm to society objection should be rejected for that reason. But if, as I am assuming here, he is wrong, then this does nothing to undermine the principle of pure restitution itself: the claim that the offender owes a debt only to his victims. It simply changes the results of applying the theory to the particular case in question: given that the offender has wrongfully harmed both the person he robbed and the other members of his community, and given that he has therefore incurred a debt to both, he should have to make restitution to both. Far from "surrendering" one of the core components of the theory, therefore, the secondary victims response to the harm to society objection involves applying the theory consistently in light of the assumption that people who break the law often harm society as a whole.

The second objection to the secondary victims response maintains that taking money from an armed robber to pay for an extra patrol officer is functionally no different from fining an armed robber. When offenders are forced to pay fines for their offenses, after all, the money goes to the government, which uses it to provide various public services. Yet, fines are uncontroversially a form of punishment. And so, on this account, the secondary victims response to the harm to society objection addresses the legitimate demands of society only by embracing punishment as a legitimate response to violations of the law. And since punishment is explicitly excluded by the theory of pure restitution, it follows that the secondary victims response is not available to a defender of the theory of pure restitution.

But this objection is also mistaken. It may be true that, from the point of view of the offender, there is little difference between paying a fine and paying a lump sum restitution settlement to the government. In either case, the offender will lose some money because of his wrongful behavior, and in either case the money will ultimately be used to benefit others. But from the point of view of the *justification* of the claim that the state has the right to compel the offender to make such payments, there is an enormous difference. In the case of the fine, the state takes money from the offender to make him suffer. That the money is then used for some beneficial purpose is merely a positive foreseeable by-product of the act. This is what makes a fine punitive and also what makes it so difficult to justify. But in the case of restitution, the state does not aim to make the offender suffer. It aims only to take from the offender what is not rightly his, foreseeing that this will harm the offender. It is this difference between the reason that the money is extracted from the offender in the two cases that best illuminates the conceptual distinction between punishment and restitution. And in terms of this difference, it is clear that what the secondary victims response justifies is restitution, not punishment.

Finally, there is what Wilkinson calls the "problem of public perception" (1996: 40). Suppose that Larry and Moe each commit armed robbery, but for some reason, Larry's crime generates much more publicity than Moe's. In this case, Larry will have wrongfully caused more psychic harm to the community than Moe will have caused, and the secondary victims response to the harm to society objection will therefore seem to entail that Larry will owe more restitution than Moe will. Yet, this seems unfair. Larry and Moe committed the same offense, and it seems wrong that such arbitrary factors as the amount of publicity generated by their acts should determine the final cost that each will have to bear.

The problem of public perception is part of a more general problem. The more general problem arises because the amount of harm that an offender's wrongful act causes is determined by many factors that are beyond his control. How much publicity his offense generates is one such

factor, but there are many more. Suppose, for example, that Larry and Moe each punch a person in the nose with precisely the same force. Larry's victim might end up suffering much greater harm than Moe's because he might have had a preexisting condition that is aggravated by the attack, because he might be treated by a less competent doctor, and so on. The problem of public perception, in short, is simply one facet of the more general problem of moral luck.

The problem of moral luck, in turn, is a deep and perplexing one. But it is not a problem for the theory of pure restitution. The theory maintains that it is morally permissible for the state to compel an offender to compensate his victims for the wrongful harms that he is *responsible* for having caused. The problem of moral luck arises because, in cases such as these, it is difficult to determine how much harm the offender is responsible for having caused. One possibility is to maintain that an offender is always responsible for all of the harm that results from his wrongful act, including harms that would have been impossible for him to foresee. A second is to maintain that an offender is always responsible only for that amount of the total harm he caused that a reasonable person could have foreseen (where, of course, there is room for different accounts of what that would be). A third is to maintain that the first standard is more appropriate in some contexts, while the second is more appropriate in others. There are other possible responses, and numerous considerations that can be marshaled for and against all of them.

But a defender of the theory of pure restitution need not be concerned with these difficult problems because they all arise when attempting to answer a question that the theory itself does not need to answer. The theory is designed to answer the question "what is it permissible to do to an offender who has been found responsible for wrongfully causing a certain amount of harm?" The theory can therefore remain neutral on the problem of how best to determine what harms offenders can most reasonably be held responsible for. If it is thought most reasonable to conclude that Larry is responsible for causing more harm than Moe, then under the theory Larry will owe more restitution than Moe. If it is thought most reasonable to conclude that Larry is responsible for causing the same amount of harm as Moe, then under the theory Larry and Moe will owe the same amount. But in either case, the implications of the theory will not be at all objectionable: if Larry is responsible for causing more harm than Moe, then it is reasonable for Larry to owe more compensation than Moe, and if Larry is responsible for causing the same amount of harm as Moe, then it is reasonable for Larry to owe the same amount of compensation as Moe. And so, the proponent of the theory need not choose between the two accounts.

Finally, it is important to recognize that if I have been mistaken, and if the problem of moral luck really is a reason to reject the theory of pure

restitution, then it must also be a reason to reject the practice of punishment.[18] If Larry and Moe each try to kill someone, and one succeeds and the other fails due to factors beyond their control (Larry's victim is rushed to a much better hospital than Moe's victim, for example), the proponent of punishment will have to determine whether or not Larry and Moe merit the same amount of punishment. There is no reason to believe that solving the problem of moral luck will be any easier in the case of punishment than it is in the case of restitution. And so, the problem of moral luck in general, and the problem of public perception in particular, cannot be used to undermine the secondary victims response to the harm to society objection without at the same time undermining the practice of punishment as well. And since the thesis of this book is threatened only by those objections to the theory of pure restitution that are not also objections to the practice of punishment, it follows that the problem of moral luck is not a problem for the thesis of this book.

5.3.4 Nonmonetary Restitution

If the restitution that may permissibly be extracted from an offender is limited to money, then little more may be said in response to the harm to society objection. But there is nothing about the theory that commits one to this limitation. Barnett describes the theory in general as holding that the offender must "mak[e] good the loss he has caused" (1977: 219), and this characterization does not specify the form in which restitution can be made. Barnett, it is true, seems to proceed as if the only permissible form of payment is money. In all the examples of restitution he provides, for example, the offender makes monetary payments to the victim, either from his wages or from the sale of property (219) or from garnishment of his future income (220). And in discussing objections to the theory of pure restitution, Barnett at times seems to equate restitution with money. He considers the objection that some offenses cause harms that cannot be "expressed in monetary terms" as an objection to the theory itself, for example (223), and presents as an objection to the theory the claim that "monetary sanctions are insufficient deterrents to crime" (226), again equating restitution in general with monetary compensation in particular. But this, too, is a mistake, and a common one.[19] While giving money to

[18] And, for that matter, a reason to reject the practice of tort law. One of the most difficult, long-standing questions in the law of torts is whether one should be held liable for all the harm that one's negligence causes or only for the amount that a reasonable person would have foreseen [see, e.g., Kionka (1999: 92–8); White (2003: 96–102)].

[19] Abel and Marsh, for example, make the same mistaken assumption throughout their 1984 book on punishment and restitution, and Golash's brief discussion of restitution in her more recent book seems to do the same (2005: 162–5). Ellin (2000: 308) notes that victim–offender reconciliation could be pursued as an alternative to monetary

primary or secondary victims is one way to restore victims to their previous level of well-being, it is not the only way. And once this is recognized, the secondary victims response to the harm to society objection can be developed even further.

Suppose, for example, that a burglar has been very successful. He has finally been apprehended and been made to compensate his victims for the harms he has wrongfully caused them. He has returned all the stolen goods or made payments sufficient to cover the costs of replacing them, has compensated people for whatever emotional or psychological distress was involved, and has paid for each of them to have a new burglar alarm installed as a means of restoring them to their previous level of security. But suppose that all of this is still not sufficient to restore his victims to their previous level of well-being. Suppose, in particular, that even with their new alarm systems, organized community watches, and extra police officers patrolling the neighborhood, his victims have been made objectively less secure by his actions than they were before merely because he is so skilled at evasion and is still free to roam the streets at night.

In cases such as this, the theory of pure restitution may be unable to do anything more if it is limited itself to monetary restitution. But this is not a problem for the theory. Rather, it is a reason for the theory to refuse to be constrained by this limit. There are many other ways that the burglar in this case could more fully restore his victims to their previous level of safety and security. He could, for example, be compelled to wear a device by which his location could be monitored by the police at all times.[20] He could be subjected to intensive supervision, such as that accompanying probation in some cases, which often includes a curfew.[21] He could simply be locked up. In other cases, an offender might be made to take an anger management course, to undergo therapy, to give up drinking, to stay

compensation, but that is the only other suggestion he offers beyond some limited form of punishment.

[20] The idea of using such devices to monitor the location of people such as mental patients and those on probation or parole has been around since the early 1960s [see Schmidt and Curtis (1987: 137)] and has become much more promising with the more recent development of global positioning system technology [for assessments of the earlier practice, see, e.g., Blomberg, Waldo, and Burcroff (1987) and Vaughn (1987), as well as Schmidt and Curtis (1987)].

[21] Intensive probationary supervision is often assigned to intermediate offenders such as nonviolent felons. Although the details of such programs vary considerably, common practices include frequent personal contact with a probation officer, drug and alcohol testing and counseling, payment of a supervision fee, and mandatory employment, as well as a curfew, all of which could be justified as forms of restitution if needed to help restore the community to its previous level of well-being. For a useful description of some of these programs, see Petersilia (1987: 15–19) and Clear, Flynn, and Shapiro (1987: 31–41). For various views on the success of these programs see, in addition to these sources, Bennett (1987), Latessa (1987), and Pearson (1987).

away from certain areas or certain people or people under a certain age, and so on.[22] If one or more of these impositions are necessary for an offender's victims to be fully restored to their former level of safety and security, then he owes it to them to undergo these impositions, and they could be fully justified by the theory of pure restitution.

As with a lump sum payment to the community, a critic might again object at this point that the theory is accommodating the legitimate grievances of the community by abandoning restitution and retreating to punishment. What, after all, is locking up an offender under such circumstances if not punishment? McCarthy, for example, characterizes home detention and intensive supervision as "humane, but punitive" on the grounds that "they deliberately impose suffering through the deprivation of liberty" (1987b: 2), and Schmidt and Curtis describe as a benefit the claim that "[h]ome incarceration meets the public demand for punishment" (1987: 142). Indeed, some writers have taken what I am presenting here as an extension of the theory of pure restitution and have tried to turn it into a novel and defensible theory of punishment. On what Holmgren defends as a "restitutive theory of punishment," for example, punishment is justified as a form of restitution: "The persons who commit crimes deprive members of the community of the security they are entitled to, and part of that security is restored to them when those criminals are punished" (1983: 45, 41).[23]

But this is a mistake, and an important one. Not every legally imposed restriction on a person's freedom of movement is punishment. It is punishment only if it is done with the aim of making an offender suffer for his offense.[24] McCarthy, for example, acknowledges this when she appeals to the claim that home detentions "deliberately impose suffering," but there is nothing intrinsic to incarceration or intensive supervision that requires that it be done to impose suffering rather than to bring about some other result, with the harms to the person being detained being foreseen rather than intended. Consider, for example, a defendant

[22] A New Mexico law, for example, could be justified along the same restitution-based lines. It requires some repeat drunken drivers to install ignition locks on their cars: "The device, at a cost of $60 to $70 per month for the offender, requires the driver to blow into a tube connected to the ignition. If the driver's blood-alcohol level is over 0.02 percent, the car will not start" [Sink (2002: A15)].

[23] See also Holmgren (1989: 143–4). The same position is defended as a debt-based theory of punishment in Hershenov (1999). I respond to this position more directly in Section 5.13.

[24] Note that Holmgren herself recognizes this. She accepts Benn's definition of punishment on which "the unpleasantness is essential to it, not an accompaniment to some other treatment" [the definition is cited in (1983: 37)]. So, on Holmgren's own terms, locking up an offender with the aim of restoring the community to its previous level of security cannot count as punishing the offender. See also Wood (1997) for a useful discussion of the way in which preventive detention need not be associated with punishment.

awaiting trial who is jailed because he is considered a risk to flee. The restriction on his freedom of movement is every bit as harmful to him as that of a prisoner serving a sentence is to him. But pretrial incarceration, even though it involves the same treatment, is not punishment because it is not done with the intent of inflicting harm. The same can be said of many other practices including curfews, quarantines, travel restrictions, restraining orders, conditions of probation or of making bail while awaiting trial, and so forth. In the kind of case I am considering here, if the burglar is locked up, this is done with the intent of restoring his victims to their previous level of security. Since this, rather than harm to the offender, is the state's aim, it does not count as punishment. And so, there is no reason that it must be ruled out by the theory of pure restitution. Indeed, if it is necessary to restore people to their original level of well-being, it may be required by the theory.[25]

Finally, it is important to emphasize that the difference between incarcerating an offender on punitive grounds and incarcerating him on restitutive grounds has crucial practical significance as well. Suppose, for example, that a particular offender is to be incarcerated, and the question is whether he should be detained in his home and monitored electronically or imprisoned. The literature on home detention strongly suggests that in many cases, detention at home decreases his chances for committing further offenses and is less expensive than imprisonment [see, e.g., Schmidt and Curtis (1987: 141–2)]. If detention is justified on the grounds of restoring the community to its previous level of well-being, these considerations would favor home detention. But it is surely much more safe, comfortable, and pleasant to be confined in one's own home than in prison; so, if incarceration is justified as punishment, these considerations would favor imprisonment. In addition, when offenders are incarcerated on punitive grounds, they are routinely deprived of goods such as cigarettes, television, exercise equipment, and a long list of other things that might make life in prison less unpleasant. If the goal of incarcerating an offender is to make him suffer, these deprivations will often be justified. But if the goal is simply to ensure that his community is restored to the level of security it enjoyed prior to his offense, then there will be no justification for making his life any less pleasant than is required by his incarceration. The use of nonmonetary forms of restitution, therefore, means neither that the theory of pure restitution must lapse into a justification of punishment at a theoretical level nor that the state is entitled to treat people in the ways that punishment treats them at a practical level. And so, the harm to society objection ultimately fails to

[25] Golash, it should be noted, does acknowledge that some forms of intensive supervision and preventive detention can be *consistent* with a theory of pure restitution, but she does not recognize that they can be *justified* by the theory as well (e.g., 2005: 160, 166).

provide a reason to reject the theory or to embrace the appeal to necessity. What the offender's harm to society renders necessary is not punishment but more, and more effective, restitution.

5.4 THE IRREPARABLE HARMS OBJECTION

A critic of the theory of pure restitution might concede at this point that restitution without punishment could work well enough for certain legal offenses, those for which it is relatively easy to measure the amount of harm that an offender has caused and thus relatively easy to determine how an offender might compensate his victim(s). But, the critic might then charge, the theory fails in the case of other, generally more serious offenses, in which the harm caused seems literally to be beyond repair. The most common such cases cited by critics of the theory are rape and murder [e.g., Dagger (1991: 32); Walker (1993: 68)].[26] In these cases, as well as those involving other kinds of physical assault, it may seem preposterous, not to mention repulsive, to suggest that an offender could ever repair the damage he has wrongfully caused. The theory of pure restitution maintains that when an offender breaks the law, the state may compel him to repair the damage he has caused but may not punish him. And this seems to entail that, if the harm cannot be repaired, then the state cannot do anything to the offender at all. An offender who causes reparable harm by stealing and destroying an expensive computer, on this account, might face serious consequences for his offense, but an offender who causes irreparable harm by committing rape or murder would face none at all. And thus, again, the critic maintains that punishment is necessary.

The irreparable harms objection rests on two claims: the claim that people who commit such offenses as rape and murder would face no (or at least no serious) consequences and the claim that this result is unacceptable. I will consider these two claims in reverse order. The second claim, that the implication identified by the irreparable harms objection warrants rejecting the theory of pure restitution, is attractive at first glance. But, upon further consideration, it must be rejected. The reason for this has to do with a very general fact about the relationship between normative principles and empirical facts. Suppose, for example, that there is a moral principle of the form "in circumstances of type *C*, it is morally permissible to do *a*, *b*, or *c* but not *d*, *e*, or *f*." Then the fact that there are sometimes cases of type *C* in which it is not possible to do *a*, *b*, or *c* does not, in itself, impugn the validity of the moral principle. It simply shows that in some cases, there is no morally permissible response. Suppose, for example, that there is a moral principle that when a patient

[26] The objection is also pressed by Hoekema (1991: 338–9) and Tunick (1992: 160).

is dying, it is permissible to save her by transplanting an organ from a willing donor, but it is not permissible to save her by stealing an organ from an unwilling donor. This principle is not invalidated by the observation that in some, perhaps even many, cases there is no willing donor. This merely shows that in some, perhaps even many, cases there is no permissible way to save the patient. This is, of course, a regrettable fact, but that is all it is. If I borrow money from you and then lose all my money (and, let us say, also lose my ability to earn more money), then I cannot pay you back. But this does nothing to impugn the principle that people should repay their loans. Nor does it provide a reason to think that you are suddenly allowed to treat me in ways that would otherwise be impermissible, say by deliberately harming me. Again, it simply shows that sometimes people can't do what they should do or what they would be permitted to do were this possible. In precisely the same way, if there are cases in which it is not possible for an offender to restore a victim to his previous level of well-being, this does nothing to invalidate the principle that the state may compel offenders to make restitution to their victims but may not punish them. It simply shows that when offenders violate the law, it is not always possible for the state to do everything that it would be morally permitted to do in response.

I believe that this first response to the irreparable harms objection is sufficient. But suppose that it is not. Suppose that the (presumed) fact that the theory of pure restitution implies that in some cases it would not be permissible for the state to do anything to lawbreakers is a good reason to reject the theory. If this is the case, then the failure of this first response will, indeed, be a problem for the proponent of the theory of pure restitution. But, it is important to emphasize, it will not be a victory for the proponent of punishment. For if pure restitution must be rejected for this reason, then so must punishment. Suppose, for example, that Larry is dying of cancer. He has only a few days to live, and the quality of his life is so poor that he is indifferent between dying now and dying in a few days. Larry maliciously and savagely kills several of his nurses. If punishing people for breaking the law is morally permissible, then it will be morally permissible for the state to severely punish Larry. But given the facts of the case, it is impossible for the state to do anything to him that will make him significantly worse off than he already is. As a result, in this case it is not possible for the state to punish an offender.[27] If pure restitution must be rejected because, in some cases, we cannot extract the restitution to which we would be entitled, then punishment would also have to be rejected because, in some cases, we cannot impose the punishment to which we would be entitled. And since I have argued that a *reductio ad*

[27] For a compelling fictional representation of a relevantly similar case that makes this point quite powerfully, see McBain (2005).

absurdum objection to pure restitution will count against the thesis of this book only if it does not also apply to punishment, this provides a second reason to reject the irreparable harms objection.

I have argued that the fact that offenders cannot repair irreparable harms does not count against the theory of pure restitution and that, if it does, then the fact that some offenders cannot receive the punishment they should suffer if punishment is justified counts equally against the practice of punishment. Finally, however, and most importantly, it should be emphasized that even if an offender cannot fully repair all the harm he has caused in the cases pointed to by proponents of the irreparable harms objection, there are many things he can do to make partial compensation. This is enough to ensure that those who commit serious offenses will face serious consequences even if (perhaps especially if) they cannot fully repair the harms they have caused. And this is enough to establish that offenses involving irreparable harm provide no reason to believe that punishment is necessary. Since proponents of the irreparable harm objection almost always appeal to the cases of rape and murder in pressing their criticisms, I will try to say something about each of them, beginning with rape.

5.4.1 The Rape Victim

Suppose that a woman is raped. I will take it as uncontroversial that the rapist's act is wrongful and is harmful to his victim. Under the theory of pure restitution, the state should compel the rapist to restore his victim to her previous level of well-being. However, this is impossible. Nothing that is done will ever allow her to return fully to the quality of life that she previously enjoyed. Even if we grant this assumption, it does not follow that under the theory the state should do nothing at all. If the victim cannot be fully restored to her previous level of well-being, then the state should compel the offender to do the best he can to bring about that result. What would that amount to? I do not claim to know. But I do not need to know in order to respond to the rape-based version of the irreparable harms objection. All I need to note is that there are experts in treating rape victims and helping them to recover from their ordeals, and that the theory would direct a court to consult such experts in order to determine what should be done to help the victim in any particular case. Perhaps this would include increasing the security of the victim's home, hiring a bodyguard for her, preventing her assailant from coming within a certain distance of her, psychological therapy, a relaxing vacation, or relocating to a new city. Perhaps it would involve many other things that I have not thought of. The specifics of what would best help a particular rape victim to recover from her ordeal do not matter. The theory itself need not answer this question in order to answer the question of what should be done with the offender. The courts should do their best to

determine what would help the victim and then compel the offender to provide it. This is not to insist that as long as such things are done, the woman will be fully restored to her previous level of well-being. It is simply to say that we can make some progress in that direction, and this is what the theory of pure restitution would require of the rapist.

A rapist, of course, does not simply harm the woman he rapes. He significantly decreases the objective level of security of those who live in his community. This suggests that the considerations identified in developing the secondary victims response in Section 5.3 would generate reasons to impose a still larger debt on him, above and beyond the debt he would owe to his victim. I will indicate the ways in which further forms of restitution could be justified by such an appeal in Section 5.4.2. The result of combining restitution to the primary victim, discussed here, with restitution to the secondary victims, discussed there, will be that a rapist would ultimately face severe consequences on the theory of pure restitution, more than enough to overturn the claim that punishing the rapist is necessary because the theory cannot do anything, or anything significant, in response to rape. Before turning to the subject of secondary victims, however, I want to say something about the problem posed by the primary victim in the case of murder. And before doing that, I want briefly to address three objections that might be raised against what has already been said about the case of rape.

The first objection maintains that there is something wrong about putting a price on the ordeal of a rape victim, that accepting such payment would trivialize what happened to her or even legitimize the rapist's act. This argument has been made, for example, about reparations paid by the German government to survivors of the Holocaust, and some survivors have refused to accept such payments for precisely this reason. It is difficult to know how to evaluate this argument. In the case of civil lawsuits, accepting monetary compensation for damage to one's property or body does not seem to trivialize or legitimize the harm. If a patient accepts compensatory damages from the doctor whose careless treatment caused her to become permanently paralyzed, for example, I doubt anyone would think that this meant that the harm she suffered was trivial or that the doctor's reckless behavior was acceptable. But perhaps the intimately personal nature of the violation that rape represents renders the acceptance of payment more dubious. Still, it seems to me that at most, this argument shows that a victim might have good reason to decline the payment, say by passing it on to a worthy charity. In the case of the Holocaust survivor, for example, it may seem plausible to think that she should not take the money, but it seems implausible to think of this as a reason to believe that the German government would not be obligated to offer it to her in the first place. And the same therefore seems true in the case of the rape victim.

A second objection to my account of the rape case arises from a distinction between the harmfulness and the wrongfulness of an act. Rape, a critic might suggest, is extremely wrong, but it need not be extremely harmful. Indeed, there are cases in which the victim is unconscious when the rape occurs. Therefore, it might be argued, in these cases the victim is greatly wronged but not hurt. This possibility points to a potential problem for the pure restitution account: punishment can track the wrongfulness of an act but restitution must track its harmfulness. And this, in turn, suggests that in a system of punishment the consequences for a rapist would be suitably serious, while in a system of pure restitution, at least in cases such as that of the unconscious rape victim, they would be objectionably meager. And so, once again, the prospect emerges that punishment is necessary in the case of rape.

This second objection rests on a confusion between being hurt and being harmed. It may well be true that a woman who is raped while unconscious is not hurt. But it is not true that she is not harmed. Nor is it true that the harm to her is trivial. Admittedly, it is difficult to provide a precise account of the nature of the harm done to the rape victim. But it is not difficult to see that great harm is done to her nonetheless. Consider, for example, the question of how great a sacrifice a parent would make to prevent a particular act from harming her child. Presumably, an increase in the sacrifice that she would be willing to make on her child's behalf is a function of the amount of harm that the act would cause the child rather than a function of the wrongness of the act itself. A parent would do much more to prevent her child from being killed by a mildly negligent action, for example, than to prevent the child from being injured by someone maliciously intending to kill him. The latter act would be more wrongful, but the former would be more harmful. With this test in mind, it seems clear that a parent would put much more weight on preventing her child from being raped than on preventing many other clearly harmful acts, such as acts that caused a loss of money or property to her child or a broken bone. This strikes me as good evidence that we do think of rape as a very serious harm to the victim, even if it does not cause any physical suffering. And all of this, of course, is independent of appeals that can be made to the likely long-term psychological suffering that will occur as a result. For all of these reasons, it seems to me reasonable to insist that even if we cannot precisely identify the harm caused by rape, and even if we cannot say precisely what would best help a victim to recover, we can say enough to conclude that the harm to the victim is serious and that the rapist's debt to his victim is therefore serious as well.

Finally, it might be objected that in responding to the question of how much restitution the rape victim is owed by deferring to experts, I am simply avoiding the hardest part of the question. But this objection seems to be misguided for two reasons. First, courts already do this when it is

difficult to determine an appropriate amount and form of restitution. Rather than trying to answer the question themselves, they often appoint a "special master" with relevant expertise. In April 2003, for example, a federal judge appointed a special master in a case in which she ruled that she would require restitution to be made by a terrorist suspect who had stabbed a corrections officer in the eye with a sharpened comb while he was in custody. The officer was nearly killed in the attack and was left with severe brain damage ["Restitution Sought in Guard's Stabbing"(2003: A29)]. In this case, it is extremely difficult to determine what would constitute a reasonable amount and form of restitution. But this fact did not lead the court to say that no restitution should be made. It simply led the court to defer to an expert in the area as a way of doing as well as possible. My defense of the theory of pure restitution on this point, therefore, consists of saying that the courts should continue to do something that they already do without controversy in the first place. The second reason for rejecting this concern about my deference to authority at this point is that if this is a problem for a defender of the theory of pure restitution, then it will surely be an equal problem for a defender of punishment. If it is difficult to determine how much restitution to extract from the terrorist suspect who stabbed the corrections officer, it is just as difficult to determine what level of punishment would deter others from stabbing corrections officers in the future, or would inflict on those who do the amount of suffering they deserve, or would maximize the prospects for morally educating them, and so on. There is nothing amiss, then, about a defender of the theory's declining to specify the form and amount of restitution that would be appropriate in a particular case of rape.

5.4.2 The Murder Victim

I have focused here on rape since this case is frequently raised by critics of the theory of pure restitution. But it should be clear that if the responses provided here are successful, they should apply to most other offenses in which we doubt that the victim can ever be fully restored. The one apparent exception to this claim, however, is murder. When the offender kills his victim, after all, the victim cannot be even partially restored to her previous level of well-being. This seems to suggest that under the theory of pure restitution, there is nothing that the state may permissibly do in response to murder – plainly an unacceptable implication. So, perhaps punishment is necessary after all, but just in the special case of murder.

Murder strikes many people as providing one of the most damning objections to the theory of pure restitution. And the few people who have attempted to defend the theory have generally done little to

overcome it.[28] There are two ways in which a murderer causes serious harms to others – directly, by taking the life of his victim, and indirectly, by reducing the objective level of security and by imposing other costs on the rest of the community – and thus two ways in which a defender of the theory can respond to the problem. When both kinds of harm are adequately accounted for, the irreparable harms objection can be overcome and punishment shown to be unnecessary even for murder.

The direct response turns on three claims, which I will refer to as the "transferability claim," the "substitutability claim," and the "pricing claim." Each claim, taken individually, is extremely difficult to deny. All of the claims, taken together, show how the theory of pure restitution can overcome the irreparable harm objection even for murder and even if we limit our focus to the harm done to the victim.

The transferability claim maintains that a debt owed by one person to another can be transferred to a third party when the person who is owed something dies before the debt is fully paid. This claim is extremely difficult to deny. Suppose, for example, that Larry hires Moe to paint his house, agrees to pay Moe $5,000 upon completion of the work, and Moe completes the work, meeting all of the agreed-upon specifications. Suppose also that Moe dies suddenly just as Larry is about to pay him. In this case, it would be implausible to insist that Larry's debt has vanished with Moe's death. Larry still owes $5,000, and if Moe is no longer around to collect it, then the money is owed to Moe's estate. This is how the law currently operates, and it is uncontroversial.

The substitutability claim maintains that if one person cannot pay another precisely what she owes him, she can still be obligated to fulfill her debt by substituting something else of comparable value. This claim, too, is difficult to deny. Suppose, for example, that Larry assaults Moe, causing one of Moe's kidneys to fail. Moe now needs a new kidney to restore him to his original level of well-being. If Moe's kidney could be fully repaired, then the theory of pure restitution would support compelling Larry to pay the costs involved. But suppose that the kidney could not be repaired. It would be absurd to suppose that this would relieve Larry of his debt to Moe. If, for example, Moe could receive an artificial kidney, then Larry would still be obligated to pay for it. Compelling an offender to provide something of comparable value when the original good cannot be returned or restored is, again, how the law already operates, and again, it is uncontroversial.

[28] Abel and Marsh, for example, say little more than that the victim's family should be entitled to compensation for their loss and that some murderers might spend the rest of their lives working in prison factories to accomplish this (1984: 161, 183–4). But this does little to explain how to handle cases where the victim leaves no survivors (as when an entire family is massacred) or to explain why any murderers would have to be imprisoned if they could earn more money for restitution without being imprisoned.

The pricing claim maintains that it is possible, at least in principle, to put a dollar value on a person's life. If Curly kills Larry, for example, the claim maintains that there is a dollar amount that would equal the value of Larry's lost life. This claim, too, is difficult to deny. To deny it would be to say that Larry's life was worth no money at all, and this is plainly unacceptable. It would imply, among other things, that it would not be worth spending any money to save Larry's life. The claim that a dollar amount can be attached to an individual's life is, once again, embedded in the law. Every time there is a wrongful death suit, for example, a court must determine the value of the life that was lost. And while there might be controversies over precisely how this should be done, the fundamental claim that it should be made, and acted on, is once again uncontroversial.

Let us now suppose that the transferability, substitutability, and pricing claims are correct. If this is so, then the theory of pure restitution can overcome the irreparable harms objection even in the case of murder and even if we limit our focus to the harm done to the murderer's primary victim. For suppose, again, that Curly has wrongfully killed Larry. By killing Larry, he has taken away his life. If it were possible for Curly to give Larry back his life, he would owe it to Larry to do so, since this would restore Larry to the condition he rightfully enjoyed prior to Curly's wrongful act. On the assumption that Larry really is dead, however, and not merely temporarily brain dead, this is impossible. But just as in the case in which Larry destroyed Moe's kidney, the fact that Curly cannot give Larry precisely what he owes him does not mean that he owes him nothing at all. According to the substitutability claim, it means that he owes him something of comparable value. And according to the pricing claim, this something of comparable value can be stated in terms of dollars. Suppose, for the sake of the example, that Larry's life was worth $25 million. In that case, the result of combining the substitutability and pricing claims is that Curly would now owe Larry $25 million. But, of course, Curly can't give Larry $25 million any more than he can give him his life back. Larry, after all, is dead. So, Curly can't give him anything. But according to the transferability claim, this does not mean that Curly owes nothing to anyone. It means that the debt he has incurred is transferred to Larry's estate. And so, the result of combining these three claims is that under the theory of pure restitution, Curly would incur a large debt as a result of killing Larry. The complaint that the theory cannot account for the claim that Curly would incur this debt is therefore false.

Two objections to the direct response to the special problem posed by the case of murder merit notice. The first objection points to cases in which the murder victim has no descendants and no will. If Larry has no rightful heirs, then there is no one to whom Curly's debt to Larry can be transferred, and thus no payment that the state will be entitled to extract

from Curly. This objection can be handled in one of two ways. When a person dies with no descendants and no will, her assets can either become the property of the state or they can become abandoned property available to be claimed by interested parties. On the former practice, if Curly's debt cannot be transferred to a descendant of Larry's or to someone named in Larry's will, then it should be transferred to the state. The state, on this account, would be entitled to extract restitution from Curly. On the latter practice, the assets that Curly would owe to Larry's descendants, if any, could, in effect, be homesteaded by the first person to initiate legal proceedings against Curly.[29] Either way, then, and regardless of the circumstances, Curly's killing Larry would impose a heavy debt on Curly.

The second objection to this position points to pairs of cases in which one person's life seems to have been worth much more than another's. Suppose, for example, that Larry kills Moe and Curly kills Shemp. Moe was young, healthy, and full of energy; Shemp was old, and sickly, and tired. If Moe's life was worth much more than Shemp's, then the response I have been offering here will seem to entail that Larry will incur a much greater debt than Curly, even though they are guilty of the same offense. And to the extent that this implication seems unacceptable, so will the direct response to the special problem raised by the case of murder.

This second objection depends on the claim that Moe's life was worth more than Shemp's. A defender of the direct response could therefore reply in one of two ways. The first would be to accept the claim and its implication. If it is true that what Curly took from Shemp was worth much less than what Larry took from Moe, then the defender of the direct response can maintain that there is nothing wrong with an outcome on which the consequences for Curly are less severe than those for Larry. If Curly destroyed Shemp's $5 poster and Larry destroyed Moe's $10,000 painting, it would seem appropriate for Larry to incur a greater debt than Curly. The second option would be to reject the claim and avoid the implication. The natural remainder of Shemp's life might mean just as much to Shemp, for example, as the natural remainder of Moe's life means to Moe. For each of them, after all, the rest of their life is all they have. Using this consideration as a measure of the loss involved, the value of Shemp's life would turn out to be equal to the value of Moe's; and thus, the amount of restitution generated by their killings would be the same in both cases. These two options are, of course, incompatible. A defender of the theory of pure restitution cannot appeal to both of them. At the same time, however, he need not choose between them. If it really is true that Larry has caused much more harm by killing Moe than Curly has caused by killing Shemp, then there is nothing amiss about burdening Larry with a bigger debt than Curly. If it is not true, then their debts will be equal.

[29] Long proposes this kind of solution (1999: 138).

But in either case, their debts will be appropriate and will thus pose no problem for the theory of pure restitution.

5.4.3 Secondary Victims

I have argued that the theory of pure restitution can justify imposing a significant debt on a rapist or murderer even if we consider only the harm that they inflict on their victim. But this direct response to the problem raised by rape and murder can be greatly strengthened by adding the fact that the rapist and the murderer also impose serious costs on the rest of society. And once these further costs are taken into account, it should become clear that the theory of pure restitution can provide adequate treatment of rape and murder even if my appeal to the costs to the victims themselves is ultimately deemed inadequate.

To begin with, it is clear that there will typically be other individuals who will be harmed by rape or murder: the victim's family, friends, coworkers, employers, employees, and so forth. In all of these cases, there would be grounds for an additional debt. More generally, and perhaps more importantly, everyone in a community is made significantly worse off by rape and murder in a number of ways. Their level of subjective anxiety is likely to go up. Their level of objective security is likely to go down. There can be negative effects on the value of their property, and so on. In short, all the points that were raised as part of the secondary victims' response to the harm to society objection in Section 5.3.3 apply here as well. Thus, it seems most likely that at least in typical cases of rape and murder, and probably in the vast majority of cases, the theory of pure restitution as I am defending it here would require far more than a monetary payment from the offender, significant as that burden might be. Something would have to be done to restore the community to its prior level of security, and this would likely involve, at the least, constant monitoring of the offender, if not simply preventive incarceration.[30] Since this is so, there is no reason to believe that a society that abolished punishment could not protect itself from rapists and murderers. We do not punish rapists and murderers who are found to be insane, after all, and this does not prevent us from protecting ourselves from them. There is no reason to think that things should be different in the case of sane offenders. If abolishing punishment meant that the state could never lock

[30] Two cases might be considered as possible exceptions: mercy killings, in which the harm to the victim and the resulting threat to the community seem minimal or nonexistent, and abusive relationship cases, where, e.g., a woman kills her abusive husband, though not strictly in self-defense. In these cases, the harm to the community may be minimal, so the consequences for the offender may be less severe. But in such cases, this outcome does not seem to be sufficiently objectionable to serve as a basis for rejecting the theory of pure restitution. Certainly it does not seem to establish the necessity of punishment.

up a rapist or murderer to prevent him from attacking others, then perhaps the case of rape or murder would show that punishment is necessary. But it doesn't mean this, and so it doesn't show this.

5.5 THE VICTIMLESS WRONGDOING OBJECTION

I have argued that the theory of pure restitution can successfully handle cases where society as a whole seems to be a victim as well as cases where the harm done seems to be irreparable. None of these cases provide a reason to think that punishment is necessary. But what about cases in which there seem to be no victims in the first place? A number of critics have claimed that the theory of pure restitution should be rejected because it cannot accommodate such cases [(e.g., Tunick (1992: 159–60)]. What should a defender of the theory say in response to this victimless wrongdoing objection? Before attempting to develop a response, it is important to distinguish between two different cases that are often lumped together under the heading of "victimless": cases where the offender's act harms no one but himself and cases where the offender's act harms no one, not even himself. The best responses to the victimless wrongdoing objection will vary between the two cases.[31]

A possible example of the first case might be smoking marijuana. Suppose that marijuana is harmful to the person who smokes it but that smoking it harms no one else. Laws forbidding the use of marijuana would therefore be purely paternalistic. What should a proponent of the theory of pure restitution say about these cases? There are two possibilities. One is to concede that if the person who smokes marijuana is not harming anyone else, then it is not permissible for the state to do anything to him. This view is likely to strike many people as reasonable and thus as hardly a concession at all. It may, of course, come very close to the view that smoking marijuana should not be illegal in the first place, but if it is true that the person who smokes marijuana harms no one but himself,

[31] Some writers have argued that there is a further case that can be used to press the victimless wrongdoing objection: the case in which, if many people violate a certain law they cause harm, but if a small number do so, they do not cause harm. If virtually everyone else pays their income taxes in full, for example, and I do not, then I break a justified law without harming anyone and so, on this account, owe no restitution [e.g., Miller (1978: 359); Dagger (1991: 30–1; 34–5)]. But this case can be answered in two ways. First, it is not true that an offender in such cases causes no harm. Even if I am the only person who underpays his income taxes, and even if by only $100, my wrongful act still causes some measurable harm to society: either it loses $100 worth of services that it would otherwise have enjoyed or it suffers a $100 increase in its debt, which in the long run produces negative consequences as well. Either way, I have caused some wrongful harm and must pay restitution as a result. Second, as noted later in the text, there are costs involved in detecting my cheating and apprehending me, and these costs, too, have to be borne by me on the theory of pure restitution.

then this view, too, is likely to strike many people as reasonable. And so, biting the bullet in cases involving offenders who violate paternalistic laws may be a satisfactory response to the victimless wrongdoing objection. If such laws are unjust, after all, then there is no need to justify the claim that the state is entitled to do anything to those who violate them. This is the only response to the objection that Barnett, for example, provides.[32]

But suppose that paternalistic laws are deemed to be just and reasonable. Does this mean that the theory of pure restitution would have to be abandoned? It would not. For if a person is guilty of violating a paternalistic law, his act does have a victim – himself. Where there is an offender and a victim there is room for restitution, and there is no reason that this cannot be so even in cases where the offender and the victim are the same person.

Suppose, for example, that when Larry uses illegal drugs in the privacy of his own home, he is not harming anyone else but he is still harming himself. Since the drugs Larry is using are illegal, and since we are assuming that the laws prohibiting their use are just and reasonable, the harm he is causing himself by using the drugs is caused wrongfully. Insofar as Larry is the victim of a wrongfully caused harm, the theory of pure restitution maintains that the state may compel the offender who is wrongfully causing the harm to compensate him for the harm. And since Larry is also the offender who is wrongfully causing the harm, the theory of pure restitution maintains that the state may require Larry to incur the costs involved in compensating the person suffering the wrongfully caused harms. Perhaps, for example, Larry qua offender could be required to ensure that Larry qua victim will complete a drug rehabilitation program and pay for the expenses involved. Similarly, a person who harms only himself by violating a (presumably just and reasonable) law against gambling might be compelled to help restore his financial security by, for example, having his money placed in a fund from which someone else supervises his spending.[33] Many people, of course, will find

[32] Long seems to endorse the same response in his defense of Barnett (1999: 138–9).

[33] It might be objected that since victims can, in general, forgive the debts owed to them by those who harm them, the offender who violates a paternalistic law could always avoid making restitution by forgiving himself. The state, for example, could impose on Larry a debt to himself as a result of his drug use, but Larry could then forgive the debt without having to go through treatment or pay any of the resulting expenses. But this objection neglects the fact that in applying the theory of pure restitution to paternalistic laws, it is necessary to separate the individual involved qua offender and qua victim. It is only in someone's capacity as a victim that he has the right to forgive a debt owed to him. And while it is plausible to suppose that Larry qua offender would want Larry qua victim to forgive the debt he owes him, it is no more plausible to suppose that Larry qua victim would want to forgive the debt owed to him than to suppose that any other victim (qua victim) would wish to do so.

such intrusive governmental proposals objectionable. But surely this will be because they object to such laws in the first place, not because they think that the laws are appropriate but that such responses to their violation are objectionable. Finally, even if it turns out that paternalistic laws are just and reasonable and that the theory of pure restitution is incapable of imposing any costs on those who violate them, this shortcoming seems far too modest to warrant the claim that punishment is therefore necessary. If punishment is needed to protect us from rapists and murderers, the appeal to necessity has a great deal of force; if it is needed only to prevent mature adults from engaging in behaviors that harm themselves but no one else, the appeal to necessity has virtually no force at all. And so, whether or not paternalistic laws are justified, they pose no real problem for the theory of pure restitution.

What about victimless wrongdoing in which the offender harms neither himself nor anyone else? In these cases, at least, it might seem clear that the theory of pure restitution cannot justify doing anything at all to the offender. And if laws prohibiting acts that harm no one at all are taken to be just and reasonable, then this might still seem to be a problem for the theory even if the theory can successfully accommodate paternalistic laws.

There are, however, two ways in which a defender of the theory of pure restitution can respond to this apparent problem. First, she can deny that there are any just and reasonable laws that criminalize victimless activities of this sort. If it is true that the offender's act harms neither himself nor anyone else, that is, the defender of the theory can concede that the government should do nothing about it. Now it might at first appear that this attempt to bite the bullet would pose a significant problem for the proponent of the theory. There are, after all, widespread and contentious debates about the appropriateness of legally forbidding such behaviors as gambling, prostitution, pornography, and drug use. It might therefore appear that a defender of the theory of pure restitution would have to mount a convincing argument for each of these cases in order to rebut the victimless wrongdoing objection. But this is not true. For while there are deep and divisive debates about these and other subjects, it does not follow that there is any real debate about whether acts that harm no one should be criminalized. For in any law against a victimless behavior about which there is any debate, those who favor the law do so, at least in part, because they deny that the behavior is truly victimless. Proponents of laws against pornography, for example, maintain that pornography is harmful to women in particular or to society in general. Those who favor prohibitions on prostitution make similar claims. Opponents of legalizing drugs, gambling, polygamy, or other apparently self-regarding vices emphasize the destructive impact that they believe such behaviors have on people's families, society in general, and so forth. The only way that purely victimless behaviors could pose a real problem for the

defender of the theory of pure restitution, then, would be if there were a behavior that a group of people wanted to criminalize while at the same time acknowledging that the behavior harmed no one. But it is extremely difficult to imagine such a behavior and the alleged grounds for prohibiting it. Even those who favor laws that might seem to be purely moralistic, such as laws criminalizing homosexual intercourse, do so at least in part by claiming that the behavior weakens family bonds and thus harms society as a whole.[34] And so, it is very difficult to find any real problem for the theory of pure restitution by considering victimless behaviors.

But suppose, that I am mistaken, and that there is a behavior that is not harmful to anyone but is justly and reasonably criminalized. Even if this is so, there is still a way for the defender of the theory of pure restitution to respond to the objection that would seem to arise from this case. For even if a given behavior is harmless, the fact that it has been criminalized means that a person who engages in it imposes costs on other people. Suppose, for example, that the act of painting your bedroom green is not harmful to anyone but, for some reason, it is justly and reasonably criminalized. A person who is convicted of having painted her bedroom green under such circumstances has, in fact, imposed costs on other people even though the act itself is harmless. It takes time and resources, for example, to apprehend and prosecute her. These are costs to society that she is responsible for because the act is illegal. Part of these costs is purely financial: the money that was spent on her case. But another part of the costs arises from the fact that the time and resources devoted to her case could have been used to provide other services had she not broken the law. These opportunity costs, too, can be used in determining the total harm that she has wrongfully caused to her community. And since under the theory of pure restitution the state may compel her to make restitution for the harms she has wrongfully caused, it follows that even if the act of painting one's bedroom green is not harmful to anyone, under the theory of pure restitution it is still permissible to extract restitution from the offender (assuming that the law against painting one's bedroom green is just and reasonable).[35] In the end, then, the victimless wrongdoing objection is impaled on the horns of a dilemma: either laws prohibiting victimless behaviors are unjustified, so it does not matter that the theory of

[34] Miller (1978: 359) suggests that laws forbidding cruelty to animals pose a problem for the theory [as does Tunick (1992: 159)], but there is no clear reason not to say that the animals themselves are victims or that society is indirectly harmed when animals are mistreated, say by being offended or upset. And, on either account, there is no clear reason to deny that at least some forms of restitution could be possible.

[35] Since these costs also arise when an offender's act harms only himself, it follows that this response can also be used to show that the state is entitled to extract restitution from offenders in those cases as well.

pure restitution implies that the state should do nothing to people who violate them, or the laws are justified, and it is not true that the theory implies that the state should do nothing to people who violate them. Either way, victimless wrongdoing, if there really is such a thing, provides no good reason to reject the theory of pure restitution and thus no good reason to think that punishment is necessary.

5.6 THE FAILED ATTEMPTS OBJECTION

A further objection to the theory of pure restitution appeals to a different case in which there seems to be no victim of an offender's wrongful act: a failed attempt to cause wrongful harm.

5.6.1 The Objection

Suppose, for example, that Larry points a loaded gun at Moe, pulls the trigger with the intent of firing it, and succeeds; Moe is seriously injured. In this case, there is a clearly identifiable victim, namely Moe, to whom on the theory of pure restitution, Larry owes compensation. But suppose instead that Larry points the gun at Moe and pulls the trigger with the same intent but for some reason fails. The gun, for example, misfires or the bullet misses Moe by a few inches. In this case, although Larry has attempted to wrongfully harm Moe, it seems clear that he has not harmed him, wrongfully or otherwise. Since the theory of pure restitution says that people owe restitution only when they have wrongfully harmed someone, it would seem that in this case Larry owes restitution neither to Moe nor to anyone else. And this, in turn, seems to imply that, according to the theory of pure restitution, the state should do nothing to people who attempt to harm others wrongfully as long as their attempts fail. This implication is disturbing, and a number of writers have argued that it is sufficiently objectionable to warrant rejecting the theory itself [e.g., Miller (1978: 359); Dagger (1991: 30, 33–4)]. I will refer to this objection as the "failed attempts objection."

5.6.2 The Biting the Bullet Response

One response to the failed attempts objection is to bite the bullet. If an attempt to wrongfully harm someone fails, a defender of the theory might acknowledge, then no harm has been done. If no harm has been done, then there is no victim. If there is no victim, then there is no need for restitution. If there is no need for restitution, then there is no problem. And if there is no problem, then cases of failed attempts pose no difficulty for the theory of pure restitution. The theory, on this account, cannot be faulted for failing to provide a solution to a problem that does not exist.

I will not argue directly against this response, although I doubt that many people will find it attractive. Pointing out that no one seems to merit restitution for failed attempts does little to dispel the thought that there is something preposterous about a legal system that forbids theft and murder but does nothing at all when people try unsuccessfully to get away with theft and murder. Indeed, if anything, in emphasizing how directly the theory seems to entail that nothing should be done in these cases, the biting the bullet response seems to highlight just how damaging the objection seems to be.

5.6.3 The Fear and Anxiety Response

A second response to the failed attempts objection appeals to the fear and anxiety that failed attempts can cause both to their intended victims and to others. Abel and Marsh attempt to avoid the failed attempts objection in this way (1984: 166). When Larry tries but fails to shoot Moe, for example, Larry has caused Moe no physical harm. But surely, once Moe realizes what Larry has done (or has attempted to do), he will suffer a great deal of trauma and distress. And Moe's suffering in these ways will, in turn, adversely affect other people. They will suffer increased anxiety from the discovery that their neighborhood is less safe than they thought. They will also lose whatever positive and productive interactions they might otherwise have had with Moe, had he not been negatively affected by his close brush with death. Since Larry's pointing the loaded gun at Moe and pulling the trigger was an illegal act, and since it was responsible for causing these various harms, it would therefore follow that on the theory of pure restitution, Larry would owe compensation for these harms. And so, the defender of the theory might conclude, it is simply false that the theory implies that nothing should be done in the case of failed attempts. The person who attempts but fails to commit an unlawful act can still be held responsible for the various other harms he causes.

This second response to the failed attempts objection is more promising than the first, but it too must ultimately be rejected. For it depends crucially on the contingent claim that the intended victim is, or becomes, aware of the failed attempt to harm him. As Dagger puts it, such a response "amounts to saying that the *discovery* of an unsuccessful attempt at crime, and not the attempt itself, is what makes the attempt criminal" (1991: 33). And that this result is unacceptable can be seen by considering cases where an attempt is made but not discovered. Suppose, for example, that Larry sneaks into Moe's room in the middle of the night while Moe is fast asleep, points the loaded gun at Moe and pulls the trigger, fails to shoot Moe because of a mechanical problem with the gun, and then quietly leaves before Moe wakes up. Neither Moe nor his neighbors are ever made aware of Larry's failed attempt to kill him, so Larry's act

causes neither physical nor psychological harm to Moe or to anyone else. Surely it seems that, at least in this case, under the theory of pure restitution the state should do nothing in response to Larry's failed attempt. Since this is likely to strike most people as unacceptable, it suggests that the second response to the failed attempts objection is unacceptable as well, and that punishment might thus prove necessary after all.

5.6.4 The Risk-Based Response

A satisfactory response to the failed attempts objection, then, must establish that a would-be offender would incur some sort of debt even when the failed attempt is not detected by the intended victim. It must provide a reason to conclude, for example, that Larry would owe restitution even when Moe sleeps through his intended shooting. Since a person can owe restitution only if his action has wrongfully harmed someone, this amounts to saying that a defender of the theory must establish that even in this case, Larry has wrongfully harmed someone. And, at least at first, it might seem that this cannot be so.

On closer inspection, however, it should become clear that even when no one is made aware of the failed attempt, a would-be offender does cause wrongful harm to his intended victim. He does so by exposing his victim to a risk of harm.[36] And once this risk-based response to the failed attempts objection is successfully unpacked, it should become clear that even in cases such as Larry's failure to shoot Moe in his sleep, the would-be offender, owes restitution. The best response to the failed attempts objection is one that establishes that the theory of pure restitution does not have the implication claimed by its critics .

The analysis needed to vindicate the risk-based response begins with the following claim: a state of affairs in which a person is subject to a higher probability of suffering significant harm is worse for that person than a state of affairs in which the person is subject to a lower probability of suffering such harm. A person who is about to walk through a mine field, for example, is worse off than a person who is about to walk through a park even if the person who is about to walk through the mine field is not aware of the risks involved. This claim is perhaps too obvious to need a defense. But if a defense is called for, a few considerations seem sufficient. One is this: suppose that your child was too young to make reasoned judgments about what was and was not in her best interest, and you wished to make the best decisions for her. When faced with the choice between her being in one state of affairs or another, that is, you would make the choice based solely on what was best for her. It seems obvious that you would

[36] Ellin (2000: 309) briefly suggests but does not develop the response to the objection that I attempt to develop here.

choose the state of affairs involving a lower probability of significant harm to her rather than one involving a higher probability. It follows from this choice that you would take the former state of affairs to be better for her than the latter. A second consideration is this: it seems reasonable to suppose that one state of affairs is worse for a person than another if he is willing to pay more to avoid its occurrence than he is willing to pay to avoid the other's occurrence. You would, presumably, be willing to pay more money to have someone prevent you from unknowingly entering a mine field than you would be willing to pay to have someone prevent you from unknowingly entering a safe park. And so, again, it seems clear that a state of affairs involving a higher probability of harm is worse for the person than a state of affairs involving a lower probability of harm.

To this foundational claim, the argument in response to the failed attempts objection adds the following claim: if an act causes a person to be in a worse state of affairs than he would otherwise be in, the act harms him. This claim, too, seems beyond reproach. Indeed, if anything, it seems simply to explicate what it means to harm someone. Finally, the argument adds the claim that an act that initiates an attempt to significantly harm someone creates a state of affairs in which the intended target is subjected to a greater probability of significant harm than he would otherwise be. When Larry points his gun at Moe and pulls the trigger, for example, he creates a state of affairs in which Moe is subjected to a greater probability of harm than he would have been otherwise. This claim, too, seems undeniable.

Each of these three claims seems extremely difficult to deny. But if all three of them are accepted, then the failed attempts objection can be defeated. Applied to the case of Larry and Moe, the resulting argument can be put as follows: Larry's act of pointing the gun at Moe and pulling the trigger creates a state of affairs in which Moe is subject to a higher probability of significant harm than he would be otherwise. A state of affairs in which Moe is subject to a higher probability of significant harm is worse for Moe than is one in which he is subject to a lower probability of harm. If an act causes Moe to be in a worse state of affairs than he would otherwise be in, the act harms Moe. Therefore, Larry's act of pointing the gun at Moe and pulling the trigger harms Moe.

Since Larry's act of pointing the gun at Moe and pulling the trigger harms Moe, and since his act is (we are presuming) unlawful, it follows that Larry's act wrongfully harms Moe. According to the theory of pure restitution, it is morally appropriate for the state to compel people to make restitution when they wrongfully harm others. It therefore follows that under the theory, it is morally appropriate for the state to compel Larry to provide restitution to Moe for having pointed the gun at him and pulled the trigger even though Moe was not shot. The mere fact that he wrongfully exposed Moe to a risk of harm is sufficient. Since the failed

attempts objection maintains that the theory of pure restitution should be rejected because it entails that Larry would owe no restitution in this case, the objection must thus be rejected.

5.6.4.1 The Potential Harm Objection

Two objections to the risk-based response merit attention. The first objection denies that failed attempts cause any harm. The second concedes that failed attempts do cause harm but denies that this can be used as the basis for a rational policy of restitution. The first objection maintains that exposing a person to a risk of harm creates potential harm for the person but no actual harm. In cases where the attempt to harm someone fails, on this account, the potential harm is not actualized and therefore no actual harm is done. I will refer to this as the "potential harm objection." In the case of Larry and Moe, for example, the proponent of the potential harm objection concedes that Larry's act of pointing the gun at Moe and pulling the trigger at t_1 did create the potential for harm to Moe at t_2, a potential that would not otherwise have existed. But since Moe was not harmed at t_2, the critic of the risk-based response insists that the existence of this potential has not made Moe any worse off than he would otherwise have been. Since Moe has not been made worse off by Larry's act, he has not been harmed by it. And since he has not been harmed by it, the risk-based response to the failed attempts objection must be rejected.

There is something right about the potential harm objection. It is true that the purported harmfulness of Larry's act at t_1 arises because the act creates the potential for harm at t_2. If the gun at t_1 had been empty, for example, or had merely been a toy, then the act of pointing it at Moe and pulling the trigger at t_1 would in no way have been bad for Moe.[37] But it does not follow that if no harm occurs at t_2, then exposure to the risk of harm at t_1 was not bad for Moe in the first place. This would follow only if the fact that no harm occurred at t_2 meant that the act at t_1 had not created a potential for harm. And clearly, it does not mean this. The fact that no harm occurred at t_2 establishes only that the act that made Moe worse off at t_1 than he would otherwise have been at t_1 did not also make Moe worse off at t_2 than he would otherwise have been at t_2. Larry's failed

[37] It does follow, on this account, that Larry would owe no restitution for intentionally pointing an empty gun or a toy gun at Moe and pulling the trigger while Moe slept, but this result seems reasonable enough to avoid causing problems for the account. What about the case in which Larry mistakenly believes he is pointing a real loaded gun at Moe when he pulls the trigger? In this case, it might well be true that under the theory Larry had caused no harm, but this does not mean that under the theory nothing could be done about Larry's behavior. Larry's act would still be evidence that he intended to harm Moe, which in turn could be evidence that he continues to harbor this intention. And this, in turn, could plausibly be used to ground various preventive but nonpunitive remedies such as denying Larry access to weapons, or to Moe, and so forth.

attempt to shoot Moe, in short, had no lasting negative effect on Moe. But this clearly fails to show that Larry's act had no negative effect on Moe at all. In particular, it fails to show that Moe was not made worse off at t_1 than he would otherwise have been at t_1. Moe was still made much worse off at t_1 because he was, at that time, subject to an increased risk of harm. And this is all that a proponent of the risk-based response to the failed attempts objection needs to establish in order to establish that people who fail to wrongfully harm others still owe restitution to their intended victims under the theory of pure restitution.

5.6.4.2 The Calculation Objection

A second objection to the risk-based response concedes that exposing people to the risk of harm is bad for them, but raises a puzzling question about how the badness of this risk could be used as the basis for a rational principle of restitution. The objection is most clearly explained by examining a pair of cases each of which involves one person exposing another to a risk of harm and the only salient differences concern the probability involved and the final outcomes. So consider, for example, a pair of cases in which Larry builds and mails a letter bomb to Moe, and Curly builds and mails a letter bomb to Shemp. Because of differences in their design and construction, Larry's bomb has a very high probability of exploding and seriously injuring Moe, while Curly's bomb has a very low probability of exploding and seriously injuring Shemp. Suppose that the probability of resulting harm is 9 out of 10 in the former case and only 1 out of 10 in the latter case. But suppose also that due to factors beyond their control, Curly's bomb explodes while Larry's doesn't, with the result that Shemp is seriously injured by Curly's bomb while Moe is not at all injured by Larry's bomb. The question that the calculation objection raises about such cases is: how should the amounts of restitution be determined? On the one hand, Larry exposed Moe to a much greater risk of serious harm than Curly exposed Shemp to. On the other hand, Shemp was injured by Curly's bomb, while Moe was not injured by Larry's bomb. Should Larry owe more restitution than Curly? Should Curly owe more restitution than Moe? Should both owe the same amount?

The question raised by the calculation objection is difficult, but the defender of the risk-based response to the failed attempts objection need not answer it. The theory of pure restitution maintains that offenders must compensate their victims for the harms they are *responsible* for having wrongfully caused. Cases such as those raised by the calculation objection raise difficult questions about which harms the offender is responsible for having caused. But, at least for the defender of the theory of pure restitution, it makes no difference how these questions are answered. If the most reasonable view turns out to be that Larry and Curly are only responsible for the risk of harm caused to their victims, then

since Larry caused a much greater risk of harm than Curly did, on the theory of pure restitution Larry owes much more restitution than Curly does. If, on the other hand, the most reasonable view turns out to be that Larry and Curly are responsible for all of the consequences of their wrongful actions, then since the ultimate consequences of Curly's wrongful acts involve much greater harm than do the ultimate consequences of Larry's actions, on the theory of pure restitution Curly owes much more restitution than Larry does. But either way, such cases provide no reason to deny that it would be morally appropriate for the state to compel Larry and Curly to compensate their victims for the harms they wrongfully caused them to suffer. And if a critic of the theory is tempted to reject this response as unacceptably evasive, she should consider that precisely the same problem arises for punishment: if punishment is morally permissible, then should Larry be punished more than Curly, or Curly more than Larry, or should they both receive the same punishment? Whatever considerations are available to answer this question for punishment will be equally satisfactory for answering the question for restitution. And so, the problem of identifying a rational policy for calculating the amount of compensation owed in such cases provides no reason to reject the risk-based response in particular, or the theory of pure restitution in general, and certainly provides no reason to believe that cases of failed attempts render punishment necessary.

5.7 THE NONHARMFUL ENDANGERMENT OBJECTION

The failed attempts objection arises when people intend to commit offenses against others but fail to do so. A closely related objection that also merits notice arises when people unintentionally but negligently expose others to a risk of wrongful harm. Consider, for example, a drunk driver who poses a potential threat to the well-being of others but who manages to make it home without hitting anyone. In this case, it again seems that someone has committed a legal offense without harming anyone, and so it again seems that under the theory of pure restitution, the state cannot take any action in response to his behavior. This objection from cases of nonharmful endangerment has been raised by a number of writers, including Miller (1978: 359–60), Dagger (1991: 30, 34), and Wilkinson (1996: 45–6).

As with the failed attempts objection, a defender of the theory of pure restitution could bite the bullet and concede that nothing should be done to drunk drivers unless they actually hit someone. And, again as with the failed attempts objection, he could appeal to the claim that drunk drivers cause fear and anxiety to others even if they don't end up hitting them. But biting the bullet is again unlikely to seem plausible to most people, and the appeal to fear and anxiety will fail in the case of drunk drivers

who are not noticed by others while they are on the road. It therefore seems to me that the best response to this objection is to modify the risk-based response to the failed attempts objection by treating all those who are potentially harmed by the drunk driver as victims of his wrongful act. As with the harm to society objection, attempting to extract precise compensation for each such victim will in general be hopelessly impractical. If a few hundred people were on the road the night that a particular drunk driver was driving, and each was exposed to a different degree of risk, depending on how long he or she was on the road, how close to the drunk driver he or she came, and so on, then it will not be feasible to calculate precisely how much wrongful risk each was exposed to, convert this into a dollar amount, and have the driver write each of them a check. But, again as with the harm to society objection, this provides no reason to abandon the theory of pure restitution. It simply means that the state would have to find a better way to compel the drunk driver to restore the other drivers to their previous level of security. Perhaps this would mean forcing the drunk driver to pay for the presence of an extra patrol car. Perhaps it would mean depriving him of his driver's license, or of his car, or compelling him to undergo treatment for his drinking problem. Provided that the treatment was justified as necessary to restore the other drivers to their previous level of well-being, rather than required to harm the drunk driver, such responses would all be allowed, and perhaps even required, by the theory of pure restitution. So, cases of nonharmful endangerment ultimately provide no more reason to reject the theory than do cases of failed attempts to commit wrongful acts. Neither provides any reason to believe that a system of pure restitution would fail to respond adequately to the offenses in question, and so neither provides any reason to believe that punishment is necessary.

5.8 THE MITIGATING EXCUSES OBJECTION

A further objection to the theory of pure restitution arises from cases involving what are usually called "mitigating excuses." Suppose, for example, that Larry punches Moe in the nose, Curly punches Shemp in the nose, and the total amount of physical pain, emotional distress, financial loss in the form of medical bills, ruined clothing, and so forth that Larry causes Moe is identical to that caused to Shemp by Curly. All else being equal, most people will agree that Larry and Curly should incur some kind of cost after having wrongfully harmed Moe and Shemp (whether in the form of punishment, restitution, or some combination of the two). And, again all else being equal, most will agree that since the harm that Larry and Curly caused was the same, the cost they bear should be the same. The theory of pure restitution has no difficulty accounting for this assessment.

But now, suppose that all is not equal in the two cases for the following reason: when Larry punched Moe, Moe had done nothing to provoke him. Moe had just been sitting on a park bench, minding his own business, when Larry walked up to him and, out of the blue, punched him. But when Curly punched Shemp, he did so only after Shemp provoked him. Curly had been sitting on another bench in the park, minding his own business, when Shemp approached him and began insulting Curly's mother, hurling obscenities and ethnic slurs at Curly, and so on, until finally, Curly could no longer restrain himself and punched Shemp. In this case, Curly, but not Larry, acted under extreme provocation. And while the fact that Curly was provoked does not mean that he had the right to punch Shemp, it does suggest that he merits less censure for having punched Shemp than Larry merits for having launched an unprovoked attack on Moe. The same seems to be true of other circumstances, such as cases in which a person commits an offense under extreme duress. This differential assessment is reflected both in the intuitive response of most people to such cases and in those relatively uncontroversial laws that allow mitigating excuses to reduce the sanctions against offenders. It seems uncontroversial, in short, that when mitigating circumstances exist, the negative consequences for the offender are to that extent reduced.

But, at least on the face of it, the theory of pure restitution seems unable to account for this widely accepted judgment. After all, whether or not a nose is punched in response to provocation or under duress, the resulting harm to the victim is the same: the same painful injury, the same costly medical bills, and so on. It would therefore seem that under the theory of pure restitution, Curly owes the same amount and kind of restitution to Shemp that Larry owes to Curly. And this implication may strike many as sufficiently objectionable to warrant rejecting the theory itself.

There are several reasons to reject the mitigating excuses objection. The first is that it seems wrong to insist that the total amount of harm done to Moe is the same regardless of the circumstances under which he is punched. The physical and financial costs may be identical, but the mental distress caused by an unprovoked attack must surely be greater.

Second, and perhaps more importantly, the cost to the secondary victims is considerably greater in an unprovoked attack than in a provoked one. Suppose, for example, that Larry and Curly had recently moved into your neighborhood and had begun to hang around the local park, where you enjoy spending much of your free time. When Larry punches Moe, this significantly decreases your objective level of safety and security. Larry is the kind of person who is disposed to punch people even if they do nothing to provoke him, so his continued presence in the park will be a significant hazard for you. But Curly's punching Shemp does not have

the same consequence. The fact that he punched Shemp in response to great provocation provides little reason to think that he would attack you (unless, of course, you taunt him in a similar manner). Since Larry's act causes much greater secondary harm, he owes much greater restitution than does Curly. And this is one further way that the theory of pure restitution can account for the intuition that offenders should incur a greater total cost when there are no mitigating circumstances for their offense.

Finally, and even more importantly, there is nothing about the theory of pure restitution that precludes a direct appeal to the notion of diminished responsibility. The theory, it is important to remember, maintains that people can be required to compensate their victims for wrongful harms for which they are *responsible*. The relevance of this feature of the theory in the context of the mitigating excuses objection can perhaps be best illustrated by a few extreme cases. First, suppose that I expertly and methodically rob your house after months of careful planning and deliberation. Here it seems appropriate to suppose, unless there is evidence to the contrary, that I am fully responsible for the harm I have wrongfully caused you. I would therefore owe you full compensation for this harm. Now suppose that a mad scientist kidnaps me and brainwashes me into robbing your house. In this case, the robbery causes you the same amount of harm as the first case. And I personally carry out the robbery, just as I did in the first case. But in this case, it is clear that I am not responsible for my actions in robbing your house. So, I would not owe you any restitution in this second case, although presumably the mad scientist would. This is not to say that you would have no right to recover your property from me; of course you would. It is merely to say that you would have no right to seek any further compensation from me. This second case would presumably resemble a case in which I was insane and so robbed your house with no ability to prevent myself from doing so. Again, you would certainly have the right to get your property back, but it would be wrong to hold me responsible for any further harm you incurred.

Given these cases, we are in a better position to consider this further response to the mitigating excuses objection. If there is any merit to the judgment that, for example, the provoked attacker merits less censure than the unprovoked attacker, this must surely be because the provoked attacker is not fully responsible for his actions. If, for example, it made sense to think of him as only 50 percent responsible for the attack, then it would be fair to compel him to pay for only one-half of the harm caused by his (provoked) attack. If it seems reasonable to view him as even less responsible for the harm that was caused, then under the theory he should be compelled to make reparations for even less of the harm. If it seems most reasonable to view him as even more responsible for the harm, then under the theory he should be compelled to make reparations

for even more of the harm. Once again, the theory of pure restitution will not answer the question of just how responsible each offender is. But, once again, this will not be a problem for the theory. On any reasonable account of responsibility, the theory will produce results that are perfectly reasonable given the reasonableness of the account of responsibility. There is no reason for the defender of the theory of pure restitution to choose between the varying accounts of responsibility. And so, once again, cases involving mitigating excuses provide no reason to reject the theory.

5.9 THE RICH OFFENDER OBJECTION

One of the most disturbing objections to the theory of pure restitution is that the theory would be too easy on wealthy offenders.[38] Consider, for example, a pair of cases in which an offender violently attacks an innocent victim, causing considerable physical and emotional damage. In one case, the offender is a typical middle-class person. In the other case, he is a multi-millionaire. Assuming that the resulting injuries are identical in the two cases, it would seem clear that under the theory of pure restitution, both offenders should be made to pay the same amount of restitution. But while paying compensation will be a tremendous hardship for the middle-class offender, it will be barely noticeable for the wealthy one. And this significant disparity in felt hardship between the two offenders seems to render the theory of pure restitution objectionable for two distinct reasons: it seems unfair that the amount an offender suffers should depend on how much money he has, and it seems unsafe to embrace a theory on which there will be little to deter wealthy offenders from breaking the law.

The rich offender objection is initially quite plausible. But in the end, it must be rejected for two reasons.[39] The first is that the objection rests on a

[38] The rich offender objection is pressed by, e.g., Koehl (1968: 50), Pilon (1978: 351), Hoekema (1991: 338–9), and Tunick (1992: 158). Even Fatic, whose trust-based version of the restorative justice position agrees with the theory of pure restitution in opposing punishment, endorses the rich offender objection as a reason to reject the restitution approach (1995: 153–4).

[39] Barnett's reply to the objection, that wealthy people "tend to place a very high subjective premium on their social standing and other sorts of reputational effects that would be severely damaged by a successful prosecution," may reduce the impact of the objection, but it seems insufficient as an overall response (1998: 182). Abel and Marsh offer a very different response to the rich offender objection. They propose that "regardless of available resources, the convicted wrongdoer should be required to work off the restitutionary debt (i.e., rich people should not be allowed to buy their way out of working it off)" and that "all work, of whatever market value, should be paid at the same rate when directed toward restitution" (1984: 185). But it is difficult to see how these requirements can be justified unless one claims that they are needed to ensure that the rich offender suffers for his offense, in which case the state extracts restitution from him as a means of

mistaken assumption: that if a rich offender and a middle-class offender commit identical offenses, then the amount of harm they cause is identical. This assumption is mistaken because the objective insecurity and subjective anxiety caused to others will be much greater if the offender is rich than if he is of average means.[40] Suppose, for example, that a person with a typical income is caught robbing my neighbor's home. He is caught and forced to make restitution to my neighbor: return his goods, and compensate him for his lost opportunity of using them while they were missing and for the stress, anxiety, and so on that he suffered. On the whole, the offender has to pay a fair amount of money. Given his limited income, this requirement is a significant burden. Since this is so, I and my neighbors can be relatively confident that he has learned his lesson. But suppose instead that a wealthy person robs my neighbor. The offender is barely worse off after compensating my neighbor than he would have been had he not been caught. And the consequences of this fact are important. When the typical offender robs my neighbor and then restores him to his previous level of well-being, the cost to me and my other neighbors is relatively small. But when the wealthy offender robs my neighbor and then restores him to his previous level of well-being, the cost to me and my other neighbors is significantly higher. Since the wealthy offender has barely been harmed by his action, he has barely been deterred from repeating it. This means that he has caused much more harm to his secondary victims – more subjective anxiety and a greater decrease in their objective level of security – than has the typical offender. Thus, he owes much greater compensation to his secondary victims than does the typical offender.[41]

Indeed, in serious cases, a rich offender might not only have to pay much more money than an ordinary offender, but might have to undergo further costly measures as well. Suppose, for example, that a typical offender and a rich offender have each been found guilty of an unprovoked assault on someone in the community. If even paying a great deal of money to every single person in the community would impose a barely noticeable cost on the wealthy offender, then something more than money would be required for him to restore the community to its previous level of security. Perhaps he would have to wear an ankle monitor at

punishing him. This is something that the theory of pure restitution cannot permit and so, convenient as these requirements would be in protecting the theory from the objection, they must be rejected.

[40] This point is noted by Wilkinson (1996: 50).

[41] This is not to say that the state may take more from the rich offender because this would do more to deter him or others from committing comparable offenses in the future. Such an appeal would clearly be inconsistent with the theory of pure restitution. Rather, it is to say that the state may take more from the rich offender because the rich offender causes more harm to society, which is clearly consistent with the theory.

all times, or pay to have a police officer follow him at all times, or perhaps nothing short of preventive incarceration would be necessary. The particular details do not matter. What matters is that once all of the relevant costs are taken into account, it can be seen that the theory of pure restitution does not entail that a wealthy person would suffer much less for a particular offense than would a typical person for the same offense. And since the theory does not have this implication, the rich offender objection must be rejected.

This first response to the rich offender objection seems to me to be sufficient. But suppose that I have been mistaken and that the theory of pure restitution is inequitable in the way that its critic maintains. Even if this is so, it provides little comfort for the sort of critic that I am concerned with in this chapter. For if the rich offender objection provides a sufficient reason to reject the theory of pure restitution, then it will also provide a sufficient reason to reject the practice of punishment. The punishment for many offenses, for example, is a fine. If a millionaire and a person of average means are each fined for committing the same infraction, the wealthy offender suffers much less than the ordinary one. And the same is often true in the case of prison sentences. A typical offender, for example, is likely to find it extremely difficult to find work after spending five years in prison. His sentence may ruin the rest of his life. But a wealthy offender will find it much easier to return to life after prison. And so in this case, too, the felt hardship for the wealthy offender will be considerably less than for the average offender. Hardships, in short, are considerably easier to bear when you have a lot of money. And since this is so regardless of whether the hardship involves forced compensation or imposed punishment, this fact cannot be used to reject the theory of pure restitution without at the same time rejecting the practice of punishment. And since, as I argued in Section 5.1, a reductio ad absurdum objection to the theory of pure restitution can have force against the thesis of this book only if it does not at the same time apply to punishment, this provides a further reason to reject the rich offender objection.

5.10 THE POOR OFFENDER OBJECTION

I have argued that the theory of pure restitution does not produce unacceptable results in the case of the rich offender, and that even if it did, this fact could not be used to help the critic who wishes to supplement restitution with punishment. What, on the other hand, about the poor offender? Here, too, critics have argued that the theory produces unacceptable results.[42] But here, too, in the end, it is the objection, rather than the theory, that must be rejected.

[42] E.g., Hoekema (1991: 338–9) and Hershenov (1999: 84n.).

The poor offender objection arises from the claim that in many instances an offender is too poor to fully compensate his victim. If Larry is poor, unemployed, and homeless when he vandalizes Moe's expensive sports car, example, he might cause $50,000 worth of damage and be unable to pay Moe more than a few dollars in restitution. In this case, it would seem to be impossible for Larry to even come close to fully compensating Moe. And so, the proponent of the objection maintains, the theory of pure restitution must be rejected.

Cases such as this one exist, of course, though often it may be possible to compel the offenders to work and earn enough to pay at least some of what they owe.[43] The question is why their existence should count as a problem for the theory of pure restitution. One possibility is that they pose a problem because they demonstrate that in some cases the state cannot do all that the theory permits it to do. The theory says that the state is permitted to extract $50,000 from Larry, but by hypothesis this will be impossible. But if this is what is supposed to render the case of the poor offender problematic, then the poor offender objection becomes merely a special case of the irreparable harms objection (in this case, the harm is irreparable not because of the nature of the harm but because of the nature of the person who causes it) and must be rejected for the same reasons given for rejecting that objection in Section 5.3: such cases are not problems for the theory because they show only that sometimes we cannot do all that we are permitted to do and, even if they were considered a reason to reject the practice of restitution without punishment, they would also have to be considered a reason to reject the practice of restitution with punishment, since there are also many cases in which the state cannot inflict as much suffering on an offender as the practice of punishment would warrant.

A second reason for supposing that people who cannot afford to make much restitution to their victims pose a problem for the theory of pure restitution arises from a concern about deterrence. Here the worry is that if a poor person knows that he won't be able to pay anything to repair the harm he wrongfully causes by, for example, vandalizing someone's car, then he will know that nothing will happen to him. And if he knows that nothing will happen to him, then there will be nothing to deter him from vandalizing the car. On this construal, the poor offender objection turns out to be, perhaps surprisingly, a variant of the rich offender objection. The worry in both cases is that the would-be

[43] For a useful summary of some of the ways in which prison labor has been productively employed to make, among other things, computer disk drives, blue jeans, and low-cost housing, and to provide such services as data entry, furniture reupholstering, and telemarketing, see Barnett (1998: 177–80).

offender will know ahead of time that merely being made to (attempt to) compensate his victims will leave him no worse off than before: the rich offender because he'll barely miss the money and the poor offender because he'll have no money to miss. And so, the worry in both cases is that, because of the offender's financial status, restitution without punishment will fail to deter him.

But if the worry raised by the poor offender objection is merely a reiteration of one of the concerns underlying the rich offender objection, then it must again be rejected, this time for the reasons given for rejecting the rich offender objection in Section 5.9. First, the objection, on this construal, overlooks the importance of secondary victims. If a typical offender vandalizes my neighbor's car and is forced to pay thousands of dollars as a result, then it is reasonable to conclude that he has learned a lesson and will be unlikely to vandalize my car. But if an offender incurs virtually no cost in compensating his primary victim (either because he is rich and barely notices the loss of a few thousand dollars or because he is poor and pays almost nothing in the first place), this itself will mean that he causes much greater harm to his secondary victims in the form of increased objective insecurity and subjective anxiety. If the poor offender cannot compensate me and my neighbors for this harm in a monetary form (say, by buying each of us a car alarm), then he could be required to compensate us in some other way, perhaps by being subject to a curfew, or to constant monitoring, or even, in extreme cases, to preventive detention.

Second, if the disparity between the impact of compulsory restitution on the poor and on others were considered a reason to reject the theory of pure restitution, it would also have to be considered a reason to reject the practice of punishment. A desperately poor person or a terminally ill person, after all, may have virtually nothing to lose, and therefore will be much less seriously harmed by a given amount of punishment than will a person who is better off or healthier (if Larry knows that he is going to die next week, for example, then how deterred will he be by the prospect of a twenty-year term in prison, especially one that could only be imposed after a full trial that will not even begin before he dies?). But this means that the very feature that the proponent of the poor offender objection appeals to in criticizing the practice of pure restitution – responding to violations of the law in ways that will not be very harmful to some would-be offenders – is also a feature of the practice of punishment. And so, if the objection based on this feature poses a problem for the former practice, then it also poses a problem for the latter. While the poor offender objection at first seems to pose a further problem for the theory of pure restitution, therefore, in the end it merely recapitulates other objections that have already been noted and rejected.

5.11 THE INSUFFICIENT DETERRENCE OBJECTION

A further objection to the theory of pure restitution maintains that a system of pure restitution would provide insufficient deterrence. Indeed, in its strongest form, the objection charges that such a system would provide not only no deterrence at all, but actually an incentive to break the law. Suppose, for example, that you walk into a store, put a candy bar in your pocket, and leave. You have stolen a candy bar and, as a result, you have wrongly made the store owner worse off than he was before. Before you stole the candy bar there were, say, twenty candy bars on his shelf; now, as a result of your theft, there are only nineteen. You have made him worse off by one candy bar. But if this is so, the critic charges, then all you have to do in order to make full restitution to him is return the candy bar. This restores him to his original level of well-being, since now he has as many candy bars as he had before you stole one of them. But being forced to make such restitution makes you no worse off than you were before you stole it. And so, according to this objection, you have no reason not to try to steal a candy bar every time you walk into the store. Suppose that nine times out of ten, the store's security guard catches you. In those cases, you will simply return the candy bar. And if, occasionally, he doesn't catch you, then you come out ahead. And so, overall, you have no reason not to try to steal the candy bar and at least some reason to try to steal it.

When put in this simple form, the insufficient deterrence objection can easily be defeated. For the objection, as stated, fails to take into account the secondary harms your wrongful acts cause. Your stealing candy bars causes the store owner to hire more security guards or to pass along his increased costs to his customers in the form of higher prices. So, either way, you cause more harm than the value of the candy bar alone. Once this is taken into account, it should become clear that if you get caught, you will have to do more than return the candy bar. You will also have to pay some additional costs to the owner, his customers, the local police department, some other proxy organization, or some combination of these. In addition, it is fully consistent with the theory to insist that you pay the costs involved in prosecuting your case [see, e.g., Barnett (1998: 234); Ellin (2000: 302)]. These could include the costs involved in processing the evidence and paying the salaries of a judge, court reporter, prosecutor, bailiff, expert witnesses, and so on, as well as compensating jurors who are forced to sit on your case for their time and lost wages. Doing all of this would leave you considerably worse off than you were before you first entered the store. And so, the full costs of your restitution should deter you from stealing the candy bar in the first place.

But this response prompts a further consideration, one that threatens to reinforce the initial thrust of the objection. If the detection rate for the offense I am considering committing is very high, the critic may concede,

then the mere prospect of being made to pay full restitution if I am caught should be sufficient to deter me once the effect of my wrongful act on secondary victims is taken into account. If, for example, I am 95 percent confident that if I break into your house and steal your $250 television set I will have to return it and also pay, say, $100 toward improving your home security system, then the prospect of having to make restitution for my offense if I am caught should be enough to deter me. If I have a 5 percent chance of coming out ahead by $250 and a 95 percent chance of coming out $100 poorer, then, rationally speaking, trying to steal your television set is a bad bet.

But suppose instead (and, the critic might add, more realistically) that the detection rate for the offense is considerably lower. Suppose that there is only a 50 percent chance that I will be caught. In this case, there is a 50 percent chance that I will come out ahead by $250 and a 50 percent chance that I will lose $100. Not only does this set of payoffs fail to deter me, it gives me an incentive to go around trying to steal people's television sets. And so, the critic concludes, if the detection rate is much lower than 100 percent, then the requirement to make full restitution to all of my victims and nothing more will still fail to deter me. Again, if anything, it may give me an incentive to break the law.

This version of the insufficient deterrence objection is indeed more powerful. But it is unsuccessful nonetheless. It overlooks the fact that the harm wrongfully caused to secondary victims is, in part, a function of the likelihood that the offender will be detected and apprehended. The more likely he is to escape detection, the greater the subjective anxiety and objective insecurity his offense will cause to other potential victims similarly situated, and thus the greater harm he will cause. Suppose that I break into your neighbor's house when he is not home, causing no damage to his house, and steal his television set. If the detection rate for this offense is extremely high, approaching 100 percent, then this robbery will cause you only a little added insecurity. Since the detection rate for this offense is so high, my having gotten away with robbery once is very unlikely to encourage me to try it again or to inspire copycat robberies. So, my robbery of your neighbor provides relatively little reason to think that someone will try to rob you. In addition, and perhaps more importantly, since the detection rate is so high, there is good reason to be confident that even if someone does try to steal your television, he will be captured and forced to return it along with payment for whatever secondary damages he may have caused you. In short, if I steal your neighbor's television set and the detection rate for this offense is very high, this provides very little reason to conclude that someone will try to steal your television set and succeed. And so, the amount of secondary damages that I would owe you and my victim's other neighbors would be relatively very small.

But now suppose, on the other hand, that the detection rate for robbery is very low. In that case, my successfully robbing your neighbor causes you much greater secondary harm. The chances that I will rob again or inspire others to copy me are significantly greater; and so, the increased risk of your being robbed is greater than it is when the detection rate is much higher. In addition, the chances of recovering your television set and receiving further compensation from the offender are much lower if the detection rate is much lower. In short, if I steal your neighbor's television set and the detection rate is very low, this robbery creates a much stronger reason to conclude that someone will try to steal your television set and succeed. And so, the amount of secondary damages that I would owe to you and to my victim's other neighbors would be much larger.

I have argued that the secondary harm an offender causes goes up as the probability that he will be apprehended goes down. This fact, in turn, can be used to undermine the insufficient deterrence objection even in its more sophisticated form. The objection, after all, is grounded in the thought that a would-be offender takes into account both the probability that he will be caught and the total cost he will incur if he is caught. Call the product of the odds of being caught and the total cost incurred if he is caught the "expected total cost to the offender." The objection maintains that as the detection rate goes down, the expected total cost will go down, so that the would-be offender will soon reach the point at which he expects to benefit (or at least not to be harmed) overall by breaking the law. The problem with the objection, then, is that it only looks at the effect of changes in detection rates on one-half of the equation. Once we notice that the cost to the offender, if caught, goes up as the detection rate goes down, we will see that there is no reason to worry that the prospect of making full restitution will fail to deter him even when the detection rate is much less than 100 percent.

None of this is to insist, of course, that there could be no detection rate low enough to lead to an insufficient level of deterrence. Consider, for example, a sophisticated form of computer fraud. Suppose that the chances of its being successfully detected are 1 in 10 million. In that case, it is quite possible that even after we take into account the total costs to society of such fraud and include the harms generated by the difficulty of detecting it, the would-be offender might determine that on the whole, the potential benefits of the fraud make it rational to run the risk of getting caught. In this case, it is true, the theory of pure restitution would fail to deter the rational would-be offender. But in this case, it must be acknowledged, punishment would also fail to deter the rational would-be offender unless the threatened punishment was so severe that it would clearly be disproportionate to the offense. The problem in such cases, then, is not the theory of pure restitution but the fact that some offenses

are very difficult to detect. For the insufficient deterrence objection to succeed, therefore, it must show that there are cases in which the prospect of full restitution would provide insufficient deterrence while the prospect of proportionate punishment would provide sufficient deterrence. But once the nature of the harms to secondary victims is taken into account, it becomes clear that proponents of the objection have failed to do this.

I have argued that the practice of pure restitution, when properly understood, provides sufficient deterrence.[44] But it is important to emphasize, in concluding this discussion, that I have not argued that pure restitution would provide maximum deterrence. If restitution provides a good amount of deterrence, for example, then restitution plus punishment presumably provides even more. But this fact provides little comfort for the critic of the theory of pure restitution. After all, it is almost always true that we could do more to deter people from breaking the law than we now do. We could, for example, double the current penalties for every criminal offense. We could install video surveillance cameras on every corner of every block and in every room of every building, including every private home. We could compel every citizen to wear a global positioning system ankle monitor all the time. We could compel every citizen to contribute a DNA sample to be stored in a national data bank, and so on. The fact that we could increase deterrence by doing these things does not mean that we should do them. And the same is true of the fact that we could increase deterrence by adding punishment to pure restitution. The fact that pure restitution would not provide maximum deterrence, therefore, does not pose a problem. And so, considerations of deterrence do not provide a reason to reject the theory of pure restitution. The fact that we could produce even more deterrence by engaging in punishment does not show that punishment is necessary. And so, considerations of deterrence fail to provide a reason to accept the appeal to necessity.

5.12 THE INSUFFICIENT REPROBATION OBJECTION

The insufficient deterrence objection maintains that a system of restitution without punishment would fail to provide one of the things that we expect a system of punishment to provide. A second objection makes the same claim but focuses on a different feature of punishment: its reprobative aspect. In particular, the insufficient reprobation objection maintains that a system of pure restitution would prevent the state from doing an important task: expressing society's condemnation of the offender. Since punishment can do this, and since restitution without punishment

[44] For additional responses to this objection, see also Hajdin (1987) and Ellin (2000: 303).

cannot, the objection maintains that punishment is necessary after all and that the theory of pure restitution must therefore be rejected.[45]

A defender of pure restitution, could respond by denying that it is important for the state to express society's condemnation of the offender or, more modestly, by denying that the expression of such disapproval is sufficiently important to justify intentional infliction of harm to accomplish it. This response, however, is not necessary. For even if one believes that it is important for the state to express society's disapproval of the offender, there is no reason to believe that this could not be done by pure restitution.

That this is so can perhaps be seen most easily by looking first at restitution within the home. Suppose, for example, that Larry's young son steals a toy from Moe's young son and that when Larry catches him, he forces his son to return the toy. Clearly, this would amount to compelling his son to restore Moe's son (at least in part) to the level of well-being that he enjoyed before his toy was stolen. And just as clearly, it seems to me, Larry's son would recognize that in forcing him to return the toy, his father was telling him that he had no right to take the toy in the first place. After all, if Larry's son was entitled to take the toy in the first place, wouldn't it follow that Larry was not entitled to force him to return it? When a child is forced to give back something he has taken, he understands that this is happening because he did something wrong.[46] And if this much can be clear even to a child, it should be even clearer to any offender who is competent enough to be held responsible for his unlawful behavior in the first place. It is difficult to imagine, for example, that an arsonist could be mentally fit enough to be held legally culpable for burning down a building without being competent enough to understand that he is being compelled to compensate the building's owner for the damage he caused because he did something that he had no right to do.[47] Indeed, for just this reason, victims typically see restitution as in part a symbolic statement about what happened to them [see, e.g., Strang (2002: 17)]. Even if one believes that it is important for the state to act in a way that expresses society's disapproval of the offender's act, therefore, there is no reason to conclude that the theory of pure restitution should be rejected.

[45] This objection is pressed by, e.g., Miller (1978: 359).

[46] Similarly, as Golash points out, when a civil court orders payment to be made by someone found to be in breach of contract, the court's action counts as a "refutation" of the claim that it is permissible to renege on one's contractual obligations (2005: 56).

[47] That restitution does convey social disapproval of the offender's act is stressed by some defenders of the theory of pure restitution [e.g., Wright (1996: 62–3)] and is also acknowledged by some critics [e.g., Dagger (1991: 37)].

5.13 THE PUNISHMENT AS RESTITUTION OBJECTION

The theory of pure restitution maintains that it is morally permissible for the state to require offenders to restore their victims to the level of well-being that they rightfully enjoyed before they were wrongfully harmed. In attempting to defend this theory from a number of objections, I have at several points relied, at least in part, on the claim that restitution need not be purely monetary. In response to the harm to society objection and the irreparable harm objection in Sections 5.3 and 5.4, for example, I argued that an offender might be required to remain physically segregated from others as a means of restoring his victims to the level of security they enjoyed prior to his offense. In this way, I have attempted to demonstrate that the theory of pure restitution can justify a much wider variety of responses to unlawful behavior than might at first be apparent, and that the theory can overcome many objections that might initially seem to undermine it.

A critic of the theory, however, might gladly accept all that I have said and then attempt to turn it against the theory in the form of one final objection: if an offender can be required to remain locked up as a means of restoring his victim to his previous level of well-being, then why can't an offender also be required to submit to a harmful treatment if this, too, is needed to restore his victim to his previous level of well-being? As was noted in Section 3.4.3, after all, it seems plausible to suppose that in many cases, an offender's wrongful act will cause his victim to desire revenge. This, in turn, will manifest itself as a desire that the offender suffer for his offense. In cases such as these, it would therefore seem that the state could make the victim feel better not simply by constraining the offender but by harming him. If the state harmed the offender in order to make the victim feel better, moreover, the harm to the offender would not merely be a foreseeable by-product of the state's action. Rather, the harm to the offender would be intended not as an end in itself, but as a means to the end of restoring the victim to his previous level of well-being. But this means that the theory of pure restitution would, in such cases, justify the claim that it is morally permissible for the state to intentionally harm the offender. And this, in turn, means that the theory would justify the claim that it is morally permissible for the state to punish the offender. This final objection, then, which I will refer to as the "punishment as restitution objection," maintains that we cannot accept the practice of pure restitution without also accepting the practice of punishment. More specifically, it maintains that the very considerations that justify compulsory victim restitution also vindicate a certain version of revenge-based retributivism. This version of retributivism was briefly identified at the end of Chapter 3, where it was set aside until the theory of

pure restitution had been explained.[48] Now that the theory has been explained, it is necessary to return to the concern raised by this appeal to revenge in the context of restitution.[49]

The concern generated by the punishment as restitution objection is perhaps best approached by means of an example. So, suppose that Larry vandalizes Moe's house: he breaks the windows, damages the roof, and spray paints obscene statements about Moe and about Moe's mother throughout the rooms. Larry is captured and convicted of the offense, and a judge is now attempting to determine the full amount of compensation to which Moe is entitled. She begins by considering the material costs that Larry wrongfully imposed on Moe: the costs of fixing the windows, repairing the roof, cleaning the walls, and so forth, along with the costs involved in, for example, having to wait at home for the repairman when Moe could have been doing something more valuable with his time. From this, she determines a monetary debt that Larry now owes to Moe. In addition, perhaps, the judge takes into account the decrease in personal security that Moe has suffered and so imposes a restraining order on Larry, prohibiting him from coming within a ten-block radius of Moe's home, as a means of restoring Moe to the level of security that he had previously enjoyed. Suppose, in short, that the judge claims to have arrived at a form of restitution that will come as close as possible to restoring Moe to the level of well-being that he enjoyed prior to Larry's wrongful act.

In a case such as this, the proponent of the punishment as restitution objection (or, what amounts to the same thing, the restitution-based version of revenge-based retributivism) can be understood as pressing the following claim: money may help to restore Moe to the level of material well-being that he enjoyed prior to Larry's wrongful act, and a restraining order against Larry may help to restore Moe to the level of security that he previously enjoyed, but Moe will not *feel* as good as he did prior to Larry's wrongful act until he has the satisfaction of knowing that Larry is suffering for his act. There is, the proponent of the objection urges, a certain peace of mind in knowing that the person who has wronged you is suffering as he deserves and, correspondingly, a certain frustration in knowing that he is not. Moe will suffer this frustration as long as Larry is not made to suffer for his offense, and Moe will enjoy this peace of mind

[48] It is also, in effect, the justification of punishment defended by Holmgren (1983, 1989) and Hershenov (1999) referred to in Section 5.3.4.

[49] Ellin (2000: 306) in effect bites the bullet on this objection, allowing that a restitution system might have a "victim's preference option" on which the victim could choose to enjoy the pleasure of seeing the offender go to jail as his form of compensation, but it is difficult to see how this could be allowed without conceding the revenge-based retributivist position and thus abandoning the theory of pure restitution.

only when Larry is made to suffer. Since Moe's mental anguish in this instance exists only because of Larry's wrongful act, Moe's anguish is a harm to Moe that Larry has wrongfully caused. Since only the deliberate infliction of suffering on Larry will relieve Moe's anguish, and since the state should compel Larry to compensate Moe for the harms that he has wrongfully caused him, the state should compel Larry to undergo the deliberate infliction of suffering. And this means that the state should punish Larry.

The punishment as restitution objection is a natural one, but it arises from an incomplete understanding of the theory of pure restitution. The theory does not maintain that the state should compel offenders to restore their victims to whatever level of well-being they previously enjoyed. It maintains only that the state should compel offenders to restore their victims to that level of well-being they previously and *rightfully* enjoyed, which means that they were entitled to be protected in their enjoyment of it by a just and reasonable law. If I steal goods from you that were rightfully yours, for example, then the theory maintains that I owe you compensation for their loss. But if I steal goods from you that you, in turn, had stolen from someone else, then even though my act is illegal and causes you to be worse off than you were before, the theory does not maintain that I owe you compensation for your loss. This distinction has been relatively unimportant in assessing many of the objections to the theory thus far considered. But it is essential to seeing what is wrong with the punishment as restitution objection.

To see that this is so, it may be useful to consider a case where the application of the distinction does not involve a desire for revenge and applying the lesson from this example to Larry's vandalizing of Moe's house. So, consider first a case in which the state has made it legal to keep a collie at home but illegal to keep a pit bull. Keeping any dog, of course, generates a risk that someone will be attacked by it, but the state determines that the risk generated by collies is acceptable, while the risk generated by pit bulls is not. Moe lives next door to Larry, and Larry does not own a dog. One day, however, Larry illegally purchases a pit bull and brings it home. This illegal act leaves Moe worse off in two ways: he is less secure, because it is dangerous to live next door to a pit bull, and he is less happy, because he despises dogs and is greatly distressed by the thought that a dog is living next door. In this case, under the theory of pure restitution, Larry's act has wrongfully harmed Moe (assuming that the law prohibiting pit bull ownership is just and reasonable) and the state should compel Larry to make restitution. However, that this does not mean that Larry must restore Moe to the *original* level of well-being that Moe enjoyed. Prior to Larry's wrongful act, Moe enjoyed both the security of not having a pit bull living next door and the peace of mind of not having any creature that he despises living next door. But while the former

component of Moe's previous level of well-being is entitled to legal protection, the latter is not. And so, while Larry would owe it to Moe to get rid of his pit bull, he would not owe it to Moe to refrain from replacing his pit bull with a collie. This, again, is a result of the fact that under the theory of pure restitution, an offender must only restore his victim to the level of well-being he *rightfully* enjoyed prior to the offender's wrongful act.

I take it as clear, then, that in the case of the pit bull, the theory of pure restitution would not require Larry to restore that component of Moe's previous level of well-being that concerned Moe's dislike of having a dog next door. But the lesson illustrated by this case can be applied to the vandalism case as a way of undermining the punishment as restitution objection. For in the case in which Larry vandalizes Moe's house, we can identify (at least) three ways in which Moe enjoyed a high level of well-being prior to Larry's wrongful act: he had a nice house that was in good shape, he enjoyed a high level of security in his property, and he was comforted by the mental tranquility that comes with the knowledge (or, at least, the belief) that bad people (and good people) generally get what they deserve. It feels good to see good people flourish and bad people get their comeuppance, and on the whole, Moe enjoyed a large amount of that good feeling. Now it should be clear that in the case of the first two components of Moe's overall well-being, Moe was not simply enjoying a certain level of well-being but was *rightfully* enjoying it. It is uncontroversial, that is, to say that the law protects people in the ownership of their (legally acquired) property and provides them with a certain level of security in their enjoyment of it. Laws forbidding theft, for example, or forbidding exposure of others to serious risks of harm (e.g., drunk driving, discharging a weapon in public) are by and large uncontroversial. So, it is clear that, with respect to these components of Moe's initial level of well-being, under the theory of pure restitution Larry should be compelled to restore Moe to the level of well-being that he enjoyed prior to Larry's wrongful act.

But in the case of the third component of Moe's overall well-being, the mental pleasure that Moe derived from feeling that people generally got the good or bad outcomes that they deserved, it should be equally clear that even though Moe was enjoying that feeling prior to Larry's wrongful act, he was not *rightfully* enjoying it. The law does not grant us the right to feel good about how others are faring any more than it grants us the right to feel good about what animals live next door to us. If bad people flourish and that makes me sad, my sadness makes me worse off, but this harm to me is not one from which I am legally entitled to be protected. It is difficult, moreover, to imagine any reasons to think that the law should protect me from it.[50] So, prior to Larry's unlawful act, Moe may have felt

[50] A critic of the theory might point out that my response to this objection depends on the claim that there should not be a legally protected right to feel good about how others are

good (or at least not bad) about how Larry's life was going for Larry, and Moe's feeling good about this may have contributed to Moe's own level of well-being, but Moe had no right to have this component of his well-being protected by law. The theory of pure restitution says that the state may compel an offender to restore his victim only to the level of well-being that he *rightfully* enjoyed prior to the offender's wrongful act. And as a result, while it is true that having an unfulfilled desire to see Larry suffer for his offense does make Moe worse off than he was before, and while it is also true that Moe's mental frustration is a direct result of Larry's unlawful act, the theory of pure restitution does not imply that Larry must compensate Moe for this loss. Just as Larry would be allowed, in the first case, to buy a collie even though this would prevent Moe from enjoying the tranquility he enjoyed prior to Larry's wrongful act, so would Larry be allowed, in the second case, to be protected from being intentionally harmed by the state even though this would prevent Moe from enjoying the tranquility he enjoyed prior to Larry's wrongful act. In both cases, Larry is allowed to leave Moe worse off than he was before because he does so only with respect to a component of Moe's well-being that has no legal right to protection. Since the theory of pure restitution does not justify the claim that Larry must make Moe feel better in this respect, it cannot be used to justify the claim that the state may intentionally harm Larry as a means of making Moe feel better in this respect. And since it cannot be used to justify this claim, it cannot be used to justify the claim that the state may punish Larry.[51] The attempt to extract a justification of

doing. It thus depends on a substantive claim about which laws should exist. If the critic can show that there should be such a legally protected right, then my response would be overturned. But this possibility is of little help to the critic: to do that, he will have to show why we have the right to intentionally harm others, and that is just what raised the problem in the first place.

[51] Even if the appeal to restitution could be used to justify punishing Larry in this case, it may also be worth noting, such a position would still fail to generate a satisfactory solution to the problem of punishment. Punishment, after all, involves punishing offenders even when their victims have no desire to see them suffer, and a solution based on the right of victims to satisfy their desire for revenge therefore could not account for the permissibility of punishment in such cases [this is one of several objections to the position raised by Klimchuk (2001: esp. 96–101)]. It is also unclear how this position could avoid the problem of punishing the innocent. If an offender harms a child and the child's parent, the child's parent might develop a desire to see the offender's child suffer, too, so that the offender would know how it feels to have his own child harmed. The punishment as restitution argument would therefore seem to entail that the offender would have to satisfy the parent's desire to see the offender's child suffer; this, in turn, would mean that the argument would justify punishing an innocent person as part of the payment of debt to the victim. One could try to argue that the victim has a right to see the offender suffer but not a right to see the offender's child suffer, but it is difficult to understand how this could be done without begging the question at issue, namely, whether or not there is a justification for making the offender suffer for his offense.

punishment from the theory of pure restitution thus fails, and with it, so does the punishment as restitution objection and the attempt to develop a restitution-based version of revenge-based retributivism.

The belief that it is morally permissible for the state to punish people for breaking the law is extremely common. In the first four chapters of this book, I argued that this belief is unfounded. The belief that we should do without punishment, at least until a satisfactory justification for it is found, is extremely uncommon. I have argued in this final chapter that this belief should be accepted. I have argued that it should be accepted by arguing against the appeal to necessity, the claim that punishment is necessary to preserve the social order, or at least to preserve the sort of social order that makes possible just relationships among its members. And I have argued against the appeal to necessity, in turn, by arguing that there is at least one alternative to punishment – the practice recommended by the theory of pure restitution – that shows it to be false. In defending the theory of pure restitution as a means of rejecting the appeal to necessity, however, it may seem that I have at times overstated the case for the theory in two important ways. It therefore seems appropriate to conclude this chapter, and this book, by emphasizing two ways in which the defense of the theory that I have advanced here, although important to this book's overall project, is meant to be a limited one.

First, by defending the theory of pure restitution from the many objections raised against it, it may seem that I have been insisting that the theory should be accepted. But this is not so. I have argued that there is no reason to reject the theory that is not also a reason to reject punishment, but perhaps there is good reason to reject both. And while I have argued that there is no reason to reject the theory that is not also a reason to reject punishment, I have not argued directly against the many competing nonpunitive alternatives defended by others. Perhaps there is good reason to prefer some alternative to punishment other than compulsory victim restitution. My claim here has simply been that the theory generates no problems that require punishment to solve them. This is a considerably narrower claim than the claim that the theory should be adopted, but it is sufficient for the purposes of this book: as long as there is at least one acceptable way that we could do without punishment, the appeal to necessity must be rejected. And if the appeal to necessity must be rejected, then our inability to solve the problem of punishment does suffice to show that we should abolish the practice of punishment.

Second, by arguing that the theory of pure restitution shows that there is at least one acceptable way to do without punishment, it may seem that I have been insisting that compulsory victim restitution can do everything that we want punishment to do. This is not so, either. Restitution can be used to repair a wrongful harm. In this respect, at least in some instances,

restitution can make it as if the harm had never happened. But restitution cannot make it as if the *wrong* had never happened. Restitution can erase the harm but it cannot erase the wrong. This is an important limitation on what victim restitution can accomplish. If punishing an offender could erase the wrong, make it as if the wrong had never happened, then a case might be made for accepting punishment despite everything I have said in this book. But punishment can no more erase a wrong than can restitution. A person can be unharmed, as it were, but he cannot be unwronged. Agreeing to do without punishment, therefore, will not permit us to do everything we would like to do when an offender violates a just and reasonable law, but it is the best that we can do. And, perhaps more importantly, it is the best that we may do.

Bibliography

Note: in cases where I cite articles as they appear reprinted in anthologies, I refer to the article by the date of its original publication but cite the page numbers from the anthology identified here (e.g., "Barnett (1977)" refers to the pagination in the second edition of Jeffrie Murphy's anthology *Punishment and Rehabilitation* of Randy Barnett's article "Restitution: A New Paradigm of Criminal Justice," which was first published in 1977). I have done this so that the order in which various contributions to a particular line of argument have been made is preserved in the presentation in the text (e.g., so that it is clear to the reader that Barnett's "The Justice of Restitution," referred to as "Barnett (1980)," was written after, and not before, "Restitution: A New Paradigm of Criminal Justice").

Abel, Charles F. and Frank H. Marsh. 1984. *Punishment and Restitution: A Restitutionary Approach to Crime and the Criminal*. Westport, CT: Greenwood Press.

Acton, H. B., ed. 1969. *The Philosophy of Punishment.* London: Macmillan.

Adler, Jacob. 1988. "The Rectificatory Theory of Punishment." *Pacific Philosophical Quarterly*, 69, pp. 255–81.

1991. *The Urgings of Conscience: A Theory of Punishment*. Philadelphia: Temple University Press.

Adler, Matthew D. 2000. "Expressive Theories of Law: A Skeptical Overview." *University of Pennsylvania Law Review*, Vol. 148, No. 5 (May), pp. 1364–501.

Alexander, Larry and Emily Sherwin. 2001. *The Rule of Rules: Morality, Rules, and the Dilemmas of Law*. Durham, NC, and London: Duke University Press.

Alexander, Lawrence. 1980. "The Doomsday Machine: Punishment, Proportionality and Prevention." *The Monist*, Vol. 63, pp. 199–227.

1986. "Consent, Punishment, and Proportionality." *Philosophy and Public Affairs*, Vol. 15, No. 2 (Spring), pp. 178–82.

1991. "Self–Defense, Punishment, and Proportionality." *Law and Philosophy*, Vol. 10, pp. 323–8.

Anas, Brittany. 2005. "DUI Death Now a Lesson." *Boulder Daily Camera*, February 10, p. 2A.

Anderson, Jami L. 1997. "Reciprocity as a Justification for Retributivism." *Criminal Justice Ethics*, Vol. 16, No. 1 (Winter/Spring), pp. 13–25.

Applebaum, Anne. 2003. *Gulag: A History*. New York: Doubleday.
Armstrong K. G. 1961. "The Retributivist Hits Back." In Acton, ed., *Philosophy of Punishment*, pp. 138–58.
Avio, K. L. 1993. "Economic, Retributive and Contractarian Conceptions of Punishment." *Law and Philosophy*, Vol. 12, pp. 249–86.
Bagaric, Mirko. 2000. "Double Punishment and Punishing Character: The Unfairness of Prior Convictions." *Criminal Justice Ethics*, Vol. 19, No. 1 (Winter/Spring), pp. 10–28.
Baier K. E. 1955. "Is Punishment Retributive?" In Acton, ed., *Philosophy of Punishment*, 1969, pp. 130–7.
Baker, Brenda M. 1992. "Penance as a Model for Punishment." *Social Theory and Practice*, Vol. 18, No. 3, pp. 311–31.
Barnett, Randy. 1977. "Restitution: A New Paradigm of Criminal Justice." In Murphy, ed., *Punishment and Rehabilitation*, second edition, 1985, pp. 211–31.
1980. "The Justice of Restitution." In Jan Narveson, ed., *Moral Issues*. Toronto and New York: Oxford University Press, 1983, pp. 140–53.
1985. "Pursuing Justice in a Free Society: Part One – Power vs. Liberty." *Criminal Justice Ethics*, Vol. 4, No. 2 (Summer/Fall), pp. 50–72.
1998. *The Structure of Liberty: Justice and the Rule of Law*. Oxford: Clarendon Press.
Barton, Charles K. B. 1999. *Getting Even: Revenge as a Form of Justice*. Chicago: Open Court.
Baylis, Charles A. 1968a. "Immorality, Crime and Treatment." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 36–48.
1968b. "Reply to Commentators." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 55–8.
Bean, Philip. 1981. *Punishment: A Philosophical and Criminological Inquiry*. Oxford: Martin Robertson.
Bedau, Hugo Adam. 1978. "Retribution and the Theory of Punishment." *Journal of Philosophy*, Vol. 75, No. 11 (November), pp. 601–20.
Benn, Stanley I. 1958. "An Approach to the Problems of Punishment" *Philosophy*, Vol. 33, No. 127, pp. 325–41.
1967. "Punishment." In Murphy, ed., *Punishment and Rehabilitation*, second edition, 1985, pp. 8–19.
Benn, Stanley I. and Richard S. Peters. 1959. "The Utilitarian Case for Deterrence." In Gerber and McAnany, eds., *Contemporary Punishment*, pp. 96–101.
Bennett, Lawrence A. 1987. "A Reassessment of Intensive Service Probation." In McCarthy, ed., *Intermediate Punishments*, pp. 113–32.
Berenson, Alex. 2004. "There's a Reason Your Mother Told You Not to Lie." *New York Times* (National Edition), March 7, Section 4, p. 14.
Bickenbach, Jerome E. 1988. "Critical Notice: R.A. Duff, *Trials and Punishments*." *Canadian Journal of Philosophy*, Vol. 18, No. 4, pp. 765–86.
Blanshard, Brand. 1968. "Retribution Revisited." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 59–81.
Blomberg, Thomas G., Gordon P. Waldo, and Lisa C. Burcroff. 1987. "Home Confinement and Electronic Surveillance." In McCarthy, ed., *Intermediate Punishments*, pp. 169–79.

Blume, Robert and Delorys Blume. 1989. "The Crime of Punishment." *The Humanist*, November/December, pp. 12–15, 44–5.
Blumoff, Theodore. 2001. "Justifying Punishment." *Canadian Journal of Law and Jurisprudence*, Vol. 14, No. 2 (July), pp. 161–211.
Bok, Sissela. 1978. *Lying*. New York: Pantheon Books.
Bradley, Gerard V. 1999. "Retribution and the Secondary Aims of Punishment." *The American Journal of Jurisprudence*, Vol. 44, pp. 105–23.
Braithwaite, John. 1989. *Crime, Shame and Reintegration*. Cambridge: Cambridge University Press.
Braithwaite, John and Philip Pettit. 1990. *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford: Clarendon Press.
Brandt, R. B. 1959. *Ethical Theory: The Problems of Normative and Critical Ethics*. New York: Prentice Hall.
1995. "Conscience (Rule) Utilitarianism and the Criminal Law." *Law and Philosophy*, Vol. 14, pp. 65–89.
Brunk, Conrad G. 1996. "Restorative Justice and Punishment." *Dialogue*, Vol. 35, pp. 593–8.
Burgh, Richard W. 1982. "Do the Guilty Deserve Punishment?" *Journal of Philosophy*, Vol. 79 (April), pp. 193–210.
Carcasole, Joseph. 2000. "Punishment: Self–Protection and Desert." *Public Affairs Quarterly*, Vol. 14, No. 3 (July), pp. 225–44.
Cederblom, Jerry. 1995. "The Retributive Liability Theory of Punishment." *Public Affairs Quarterly*, Vol. 9, No. 4 (October), pp. 305–15.
Champlin, T. S. 1976. "Punishment without Offence." *American Philosophical Quarterly*, Vol. 13, No. 1 (January), pp. 85–7.
Charvet, John. 1966. "Criticism and Punishment." *Mind*, Vol. 75, pp. 573–9.
Clark, Michael. 1997. "The Sanctions of the Criminal Law." *Proceedings of the Aristotelian Society*, Vol. 97, pp. 25–39.
Clear, Todd R., Suzanne Flynn, and Carol Shapiro. 1987. "Intensive Supervision in Probation: A Comparison of Three Projects." In McCarthy, ed., *Intermediate Punishments*, pp. 31–50.
Corlett, J. Angelo. 2001. *Responsibility and Punishment*. Boston: Kluwer Academic.
Cottingham, John. 1979. "Varieties of Retribution." *Philosophical Quarterly*, Vol. 29, No. 116, pp. 238–46.
Cragg, Wesley. 1992. *The Practice of Punishment: Towards a Theory of Restorative Justice*. London and New York: Routledge.
Dagger, Richard. 1991. "Restitution: Pure or Punitive?" *Criminal Justice Ethics*, Vol. 10, No. 2 (Summer/Fall), pp. 29–39.
1993. "Playing Fair with Punishment." *Ethics*, Vol. 103, No. 3, pp. 473–88.
Davis, Michael. 1983. "How to Make the Punishment Fit the Crime." *Ethics*, Vol. 93, No. 4, pp. 726–52.
1985. "Just Deserts for Recidivists." In Davis, ed., *To Make the Punishment Fit the Crime: Essays in the Theory of Criminal Justice*. Boulder, CO: Westview Press, 1992, pp. 121–48.
1986a. "Harm and Retribution." *Philosophy and Public Affairs*, Vol. 15, pp. 236–66.
1986b. "Why Attempts Deserve Less Punishment Than Complete Crimes." In Davis, ed., *To Make the Punishment Fit the Crime*, 1992, pp. 101–20.

1987. "Strict Liability: Deserved Punishment for Faultless Conduct." In Davis, ed., *To Make the Punishment Fit the Crime*, 1992, pp. 149–78.
1988. "The Relative Independence of Punishment Theory." In Davis, ed., *To Make the Punishment Fit the Crime*, 1992, pp. 18–41.
1990a. "Recent Work in Punishment Theory." *Public Affairs Quarterly*, Vol. 4, No. 3 (July), pp. 217–32.
1990b. "Using the Market to Measure Deserved Punishment: A Final Defense." In Davis, ed., *To Make the Punishment Fit the Crime*, 1992, pp. 234–53.
1991a. "Punishment as Language: Misleading Analogy for Desert Theorists." *Law and Philosophy*, Vol. 10, pp. 311–22.
1991b. "Criminal Desert, Harm, and Fairness." In Davis, ed., *To Make the Punishment Fit the Crime*, 1992, pp. 213–33.
1993. "Criminal Desert and Unfair Advantage: What's the Connection?" *Law and Philosophy*, Vol. 12, pp. 133–56.
1996. "How Much Punishment Does a Bad Samaritan Deserve?" *Law and Philosophy*, Vol. 15, pp. 93–116.
Day, J. P. 1978. "Retributive Punishment." *Mind*, Vol. 87, pp. 498–516.
Deigh, John. 1984. "On the Right to Be Punished: Some Doubts." In *Ethics*, Vol. 94 (January), pp. 191–211.
del Vecchio, Giorgio. 1965. "The Struggle against Crime." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 197–203.
Dimock, Susan. 1997. "Retributivism and Trust." *Law and Philosophy*, Vol. 16, pp. 37–62.
Dolinko, David. 1991. "Some Thoughts about Retributivism." *Ethics*, Vol. 101, No. 3, pp. 537–59.
1994. "Mismeasuring 'Unfair Advantage': A Response to Michael Davis." *Law and Philosophy*, Vol. 13, pp. 493–524.
1997. "Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment." *Law and Philosophy*, Vol. 16, No. 5, pp. 507–28.
1999. "Morris on Paternalism and Punishment." *Law and Philosophy*, Vol. 18, No. 4, pp. 345–61.
Doyle, James F. 1967. "Justice and Legal Punishment." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 159–71.
Ducasse, C. J. 1968a. "Philosophy and Wisdom in Punishment and Reward." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 3–19.
1968b. "Reply to Comments." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 30–5.
Duff, R. A. 1986. *Trials and Punishments*. Cambridge: Cambridge University Press.
1988. "A Reply to Bickenbach." *Canadian Journal of Philosophy*, Vol. 18, No. 4, pp. 787–94.
1990. "Auctions, Lotteries, and the Punishment of Attempts." *Law and Philosophy*, Vol. 9, pp. 1–37.
1993. "Review of Jacob Adler, *The Urgings of Conscience: A Theory of Punishment*," *Ethics*, Vol. 104, No. 1, pp. 181–2.
1999. "Punishment, Communication, and Community." In Matravers, ed., *Punishment and Political Theory*, pp. 48–69.

2001. *Punishment, Communication and Community.* Oxford: Oxford University Press.
2002. "Crime, Prohibition, and Punishment." *Journal of Applied Philosophy*, Vol. 19, No. 2, pp. 97–108.
Duff, R.A. and Garland, D. 1994a. *A Reader an Punishment*. Oxford: Oxford University Press.
1994b. "Introduction: Thinking about Punishment." In Duff and Garland, eds., *A Reader on Punishment*. pp. 1–43.
Ellin, Joseph. 2000. "Restitutionism Defended." *The Journal of Value Inquiry*, Vol. 34, pp. 299–317.
Ellis, Anthony. 1995. "Recent Work on Punishment," *The Philosophical Quarterly*, Vol. 45, No. 179, pp. 225–33.
1997. "Punishment and the Principle of Fair Play." *Utilitas*, Vol. 9, No. 1 (March), pp. 81–97.
2003. "A Deterrence Theory of Punishment." *The Philosophical Quarterly*, Vol. 53, No. 212, pp. 337–51.
Ewing, A.C. 1929. *The Morality of Punishment*. London: Kegan Paul, Trench, Trubner & Co.
1963. "Armstrong on the Retributive Theory." *Mind*, Vol. 72, No. 285 (January), pp. 121–4.
Ezorsky, Gertrude. 1972. "Retributive Justice." *Canadian Journal of Philosophy*, Vol. 1, No. 3, pp. 365–8.
Farrell, Daniel M. 1988. "Punishment without the State." *Nous*, Vol. 22, pp. 437–53.
1990. "The Justification of Deterrent Violence." *Ethics*, Vol. 100, No. 2, pp. 301–17.
1995. "Deterrence and the Just Distribution of Harm." *Social Philosophy and Policy*, Vol. 12, No. 12, pp. 220–40.
Fatic, Aleksandar. 1995. *Punishment and Restorative Crime-Handling: A Social Theory of Trust*. Brookfield, VT: Ashgate.
Feinberg, Joel. 1970. "The Expressive Function of Punishment." In Duff and Garland, eds., *A Reader on Punishment*, pp. 71–91.
1978. "Voluntary Euthanasia and the Inalienable Right to Life." *Philosophy and Public Affairs*, Vol. 7, pp. 93–123.
Fingarette, Herbert. 1978, "Punishment and Suffering," *Proceedings of the American Philosophical Association*, Vol. 50, pp. 499–525.
Finnis, John. 1972. "The Restoration of Retribution." *Analysis*, Vol. 32, No. 4, pp. 131–5.
1980. *Natural Law and Natural Rights*. Oxford: Oxford University Press.
1999. "Retribution: Punishment's Formative Aim." *The American Journal of Jurisprudence*, Vol. 44, pp. 91–103.
Flew, Antony. 1954. "The Justification of Punishment." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 83–102.
Fouche, Fidela. 1994. "The Irrelevance of Theories of Punishment: Creating Relevance." *South African Journal of Philosophy*, Vol. 13, No. 2, pp. 50–7.
Gaffney, Paul. 1994. "The Morality of Legal Retribution." *Contemporary Philosophy*, Vol. 16, No. 5, pp. 11–17.

Gahringer, Robert E. 1960. "Punishment as Language." *Ethics*, Vol. 71, No. 1, pp. 46–8.
1969. "Punishment and Responsibility." *The Journal of Philosophy*, Vol. 66, No. 10, pp. 291–306.
Garcia, J. L. A. 1986. "Two Concepts of Desert." *Law and Philosophy*, Vol. 5, pp. 219–35.
1989. "Deserved Punishment." *Law and Philosophy*, Vol. 8, No. 2, pp. 263–77.
1990. "The Rights–Interpretation of Desert." *Dialogos*, Vol. 56, pp. 143–56.
Gavison, Ruth. 1991. "Punishment and Theories of Justice: Comments on Sadurski and Hoekema." *Israel Law Review*, Vol. 25, Nos. 3–4, pp. 351–9.
Gendin, Sidney. 1967. "The Meaning of 'Punishment.'" *Philosophy and Phenomenological Research*, Vol. 28, No. 2. (December), pp. 235–40.
1970. "A Plausible Theory of Retribution." *Journal of Value Inquiry*, Vol. 5 (Winter), pp. 1–16.
Gerber, Rudolph J. and Patrick D. McAnany, eds. 1972. *Contemporary Punishment: Views, Explanations, and Justifications*. Notre Dame, IN: University of Notre Dame press.
Gerstein, Robert S. 1974. "Capital Punishment – 'Cruel and Unusual?': A Retributivist Response." *Ethics*, Vol. 85, No. 1, pp. 75–9.
Gert, Bernard. 1989. "Review of George Sher, *Desert*." *Ethics*, Vol. 99, No. 2 (January), pp. 426–8.
Gert, Heather J., Linda Radzik, and Michael Hand. 2004. "Hampton on the Expressive Power of Punishment." *Journal of Social Philosophy*, Vol. 35, No. 1, pp. 79–90.
Gewirth, Alan. 1978. *Reason and Morality*. Chicago: University of Chicago Press.
Golash, Deirdre. 1994. "The Retributivist Paradox." *Analysis*, Vol. 54, pp. 72–8.
2005. *The Case Against Punishment: Retribution, Crime Prevention, and the Law*. New York and London: New York University Press.
Goldman, Alan H. 1979. "The Paradox of Punishment." In Simmons et al., eds., *Punishment*, 1994, pp. 30–46.
Grupp, Stenley E., ed. 1971. *Theories of Punishment*. Bloomington and London: Indiana University Press.
Hajdin, Mane. 1987. "Criminals as Gamblers: A Modified Theory of Pure Restitution." *Dialogue*, Vol. 26, pp. 77–86.
Haksar, Vinit. 1986. "Excuses and Voluntary Conduct." *Ethics*, Vol. 96, No. 2, pp. 317–29.
Hampton, Jean. 1984. "The Moral Education Theory of Punishment." In Simmons et al., eds., *Punishment*, 1994, pp. 112–42.
1988a. "Forgiveness and Christianity." In Murphy and Hampton, *Forgiveness and Mercy*, pp. 10–13.
1988b. "Forgiveness, Resentment and Hatred." In Murphy and Hampton, *Forgiveness and Mercy*, pp. 35–87.
1988c. "The Retributive Idea." In Murphy and Hampton, *Forgiveness and Mercy*, pp. 111–61.
1998. "Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law." *Canadian Journal of Law and Jurisprudence*, Vol. 11, No. 1, pp. 23–45.

Hare, R. M. 1986. "Punishment and Retributive Justice." *Philosophical Topics*, Vol. 14, No. 2 (Fall), pp. 211–23.

Hart, H. L. A. 1959. "Prolegomenon to the Principles of Punishment." In Hart, *Punishment and Responsibility*, 1968, pp. 1–27.

1962. "Punishment and the Elimination of Responsibility." In Hart, *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Clarendon Press, pp. 158–85.

1968. "Notes." In Hart, *Punishment and Responsibility*, pp. 238–67.

Hart, Henry M., Jr. 1958. "Criminal Punishment as Public Condemnation." In Gerber and McAnany, eds., *Contemporary Punishment*, 1972, pp. 12–15.

Hawkins, D. J. B. 1944. "Punishment and Moral Responsibility." In Grupp, ed., *Theories of Punishment*, 1971, pp. 13–18.

Hawkins, Gordon. 1969. "Punishment as a Moral Educator." In Gerber and McAnany, eds., *Contemporary Punishment*, 1972, pp. 120–8.

Hershenov, David B. 1999. "Restitution and Revenge." *The Journal of Philosophy*, Vol. XCVI, No. 2 (February), pp. 79–94.

Hill, Jr., Thomas E. 1992. "Kantian Pluralism." In Hill, *Respect, Pluralism, and Justice*, 2000, pp. 11–32.

1997. "Kant on Punishment: A Coherent Mix of Deterrence and Retribution?" In Hill, ed., *Respect, Pluralism, and Justice: Kantian Perspectives*: Oxford: Clarendon Press, 2000, pp. 173–99.

1998. "Punishment, Conscience, and Moral Worth." In Hill, *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Clarendon Press, 2002, pp. 340–61.

1999. "Kant on Wrongdoing, Desert, and Punishment." *Law and Philosophy*, Vol. 18, No. 4, pp. 407–41.

Hobson, Peter. 1986. "The Compatibility of Punishment and Moral Education." *Journal of Moral Education*, Vol. 15, No. 3, pp. 221–8.

Hoekema, David. 1991. "Trust and Obey: Toward a New Theory of Punishment." *Israel Law Review*, Vol. 25, Nos. 3–4, pp. 332–50.

Holland, Gina. 2003. "States Can Post Sex Offenders on the Net." *Boulder Daily Camera*, March 6, pp. 7A, 10A.

Holmgren, Margaret R. 1983. "Punishment as Restitution." *Criminal Justice Ethics*, Vol. 2, No. 1 (Winter/Spring), pp. 36–49.

1989. "The Backward–Looking Component of Weak Retributivism." *Journal of Value Inquiry*, Vol. 23, No. 2, pp. 135–46.

Hooker, Brad. 2000. *Ideal Code, Real World: A Rule–Consequentialist Theory of Morality*. Oxford: Oxford University Press.

Hurka, Thomas. 1982. "Rights and Capital Punishment." *Dialogue*, Vol. 21, No. 4, pp. 647–60.

Husak, Douglas N. 1990. "'Already Punished Enough'." In *Philosophical Topics*, Vol. 18, No. 1 (Spring), pp. 79–99.

1992. "Why Punish the Deserving?" *Nous*, Vol. 26, No. 4, pp. 447–64.

Johnson, Oliver A. 1985. "Retributive Punishment and Humbling the Will." *The Journal of Value Inquiry*, Vol. 19, pp. 155–61.

Kasachkoff, Tziporah. 1973. "The Criteria of Punishment: Some Neglected Considerations." *Canadian Journal of Philosophy*, Vol. 2, No. 3, pp. 363–77.

Kemp, Kenneth W. 1996. "Punishment as Just Cause for War." *Public Affairs Quarterly*, Vol. 10, No. 4, pp. 335–53.
Kershnar, Stephen. 1995. "The Justification of Deserved Punishment Via General Moral Principles." *Southern Journal of Philosophy*, Vol. 33, No. 4, pp. 461–84.
1997a. "George Sher's Theory of Deserved Punishment, and the Victimized Wrongdoer." *Social Theory and Practice*, Vol. 23, No. 1, pp. 75–91.
1997b. "The Basis of Deserved Punishment Is a Culpable Wrongdoing." *Jahrbuch für Recht und Ethik*, Vol. 5, pp. 497–513.
2000. "A Defense of Retributivism." *International Journal of Applied Philosophy*, Vol. 14, No. 1, pp. 97–117.
2001. *Desert, Retribution, and Torture*. Lanham, MD: University Press of America.
2002. "The Structure of Rights Forfeiture in the Context of Culpable Wrongdoing." *Philosophia*, Vol. 29, pp. 57–88.
Khatchadourian, Haig. 1982. "Just Ethical Punishment." *Journal of Value Inquiry*, Vol. 16, No. 3, pp. 3–20.
Kidder, Joel. 1982. "A Sketch of an Integrative Theory of Punishment." *American Philosophical Quarterly*, Vol. 19, No. 2, pp. 197–202.
Kionka, Edward J. 1999. *Torts in a Nutshell*, third edition. St. Paul, MN.: West Group.
Kleinberg, Stanley S. 1980. "Criminal Justice and Private Enterprise." *Ethics*, Vol. 90, No. 2. (January), pp. 270–82.
Kleinig, John. 1973. *Punishment and Desert*. The Hague: Martinus Nijhoff.
Klimchuk, Dennis. 2001. "Retribution, Restitution and Revenge." *Law and Philosophy*, Vol. 20, pp. 81–101.
Knowles, Dudley. 1999. "Punishment and Rights." In Matravers, ed., *Punishment and Political Theory*, pp. 28–47.
Koehl, Richard A. 1968. "Professor Baylis and the Concept of Treatment." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 49–52.
Korman, Dan. 2003. "The Failure of Trust–Based Retributivism." *Law and Philosophy*, Vol. 22, pp. 561–76.
Lacey, Nicola. 1988. *State Punishment: Political Principles and Community Values*. London and New York: Routledge.
Latessa, Edward J. 1987. "The Effectiveness of Intensive Supervision with High Risk Probationers." In McCarthy, ed., *Intermediate Punishments*, pp. 99–112.
Lemos, Ramon M. 1977. "A Defense of Retributivism." *Southern Journal of Philosophy*, Vol. 15, No. 1, pp. 53–66.
Lewis C. S. 1949. "The Humanitarian Theory of Punishment." In Grupp, ed., *Theories of Punishment*, 1971, pp. 301–8.
Lippke, Richard L. 1998a. "Prison Labor: Its Control, Facilitation, and Terms." *Law and Philosophy*, Vol. 17, pp. 533–57.
1998b. "Arguing Against Inhumane and Degrading Punishment." *Criminal Justice Ethics*, Vol. 17, No. 1 (Winter/Spring), pp. 29–41.
2001. "Criminal Offenders and Right Forfeiture." *Journal of Social Philosophy*, Vol. 32 (Spring), pp. 78–89.
2003a. "Retribution and Incarceration." *Public Affairs Quarterly*, Vol. 17, No. 1, pp. 29–48.

2003b. "Victim–Centered Retributivism." *Pacific Philosophical Quarterly*, Vol. 84, pp. 127–45.
Locke, Don. 1963. "The Many Faces of Punishment." *Mind*, Vol. 72, No. 288. (October), pp. 568–72.
Loewy, Arnold H. 2000. *Criminal Law in a Nutshell*, third edition. St. Paul, MN: West Group.
Long, Roderick T. 1999. "The Irrelevance of Responsibility." *Social Philosophy and Policy*, Vol. 30, pp. 118–45.
Luke, Andrew. 1996. "Tackling Crime by Other Means." *Journal of Applied Philosophy*, Vol. 13, No. 2, pp. 179–88.
Lyons, William. 1974. "Deterrent Theory and Punishment of the Innocent," *Ethics*, Vol. 84, No. 4, pp. 346–8.
Mabbott J. D. 1939. "Punishment." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 39–54.
Mackie, J. L. 1982. "Morality and the Retributive Emotions." *Criminal Justice Ethics*, Vol. 1, No. 1, pp. 3–10.
Madden, Edward H., Rollo Handy, and Marvin Farber, eds. 1968. *Philosophical Perspectives on Punishment*. Springfield, IL: Charles C. Thomas.
"Man Asks for More Jail Time to Honor Bird." 2005. Boulder *Daily Camera*, October 21, p. 2C.
Manser, A. R. 1962. "It Serves You Right." *Philosophy*, Vol. 37, No. 142, pp. 293–306.
Marshall, James D. 1984. "Punishment and Moral Education." *Journal of Moral Education*, Vol. 13, No. 2, pp. 83–9.
Martin, Rex. 1970. "On the Logic of Justifying Legal Punishment." *American Philosophical Quarterly*, Vol. 7, No. 3 (July), pp. 253–9.
1987. "Modes of Punishment." *Public Affairs Quarterly*, Vol. 1, No. 4 (October), pp. 73–85.
1990. "Treatment and Rehabilitation as a Mode of Punishment." *Philosophical Topics*, Vol. 18, No. 1 (Spring), pp. 101–22.
Mason, Elinor. 2002. "Against Blameless Wrongdoing." *Ethical Theory and Moral Practice*, Vol. 5, pp. 287–303.
Matravers, Matt, ed. 1999. *Punishment and Political Theory*. Oxford and Portland, OR: Hart.
2000. *Justice and Punishment: The Rationale of Coercion*. Oxford: Oxford University Press.
McBain, Ed. 2005. *Fiddlers*. Orlando, FL: Harcourt, Inc.
McCarthy, Belinda R., ed. 1987a. *Intermediate Punishments: Intensive Supervision, Home Confinement and Electronic Surveillance*. Monsey, NY: Criminal Justice Press.
1987b. "Introduction." In McCarthy, ed., *Intermediate Punishments*, pp. 1–12.
McCloskey, H. J. 1954. "The Complexity of the Concepts of Punishment." *Philosophy*, Vol. 37, pp. 308–25.
1957. "An Examination of Restricted Utilitarianism" *Philosophical Review*, Vol. 66, No. 4, (October), pp. 466–85.
1963. "A Note on Utilitarian Punishment." *Mind*, New Series, Vol. 72, No. 288, p. 599.

1967. "Utilitarian and Retributive Punishment." *Journal of Philosophy*, Vol. 64, No. 3, pp. 91–110.
1978. "Crime and Punishment: Deviance and Corrective Social Therapy." *American Philosophical Quarterly*, Vol. 15, No. 2, pp. 91–8.
McDermott, Daniel. 1999. "The Duty to Punish and Legitimate Government." *The Journal of Political Philosophy*, Vol. 7, No. 2, pp. 147–71.
2001. "The Permissibility of Punishment." *Law and Philosophy*, Vol. 20, pp. 403–32.
McKerlie, Dennis. 1983. "Rights and Punishment." *Dialogue*, Vol. 22, pp. 535–8.
McTaggart, J. Ellis. 1896. "Hegel's Theory of Punishment." *International Journal of Ethics*, Vol. 6, No. 4, pp. 479–502.
Menninger, Karl. 1968. *The Crime of Punishment*. New York: Viking Press.
Metz, Thaddeus. 2000. "Censure Theory and Intuitions about Punishment." *Law and Philosophy*, Vol. 19, pp. 491–512.
Miller, David. 2003. "Comparative and Noncomparative Desert." In Olsaretti, ed., *Desert and Justice*, 2003, pp. 25–44.
Miller, Franklin G. 1978. "Restitution and Punishment: A Reply to Barnett." *Ethics*, Vol. 88, No. 4 (July), pp. 358–60.
Miller, Vanessa. 2005. "County Imposes Vigil Fees." *Boulder Daily Camera*, August 26, p. 3A.
Miller, W. A. 1966. "Mr. Quinton on 'An Odd Sort of Right'." *Philosophy*, Vol. 41, pp. 258–9.
1970. "A Theory of Punishment." *Philosophy*, Vol. 45, pp. 307–16.
Montague, Phillip. 1984. "Rights and Duties of Compensation." *Philosophy and Public Affairs*, Vol. 13, No. 1, pp. 79–88.
1995. *Punishment as Societal–Defense*. Lanham, MD: Rowman & Littlefield.
2002. "Recent Approaches to Justifying Punishment." *Philosophia*, Vol. 31, pp. 1–34.
Moore, David B. 1993. "Shame, Forgiveness, and Juvenile Justice." *Criminal Justice Ethics*, Vol. 12, No. 1 (Winter/Spring), pp. 3–25.
Moore, Michael S. 1987. "The Moral Worth of Retribution." In Murphy, ed., *Punishment and Rehabilitation*, third edition, pp. 94–130
1993. "Justifying Retributivism." *Israel Law Review*, Vol. 27, pp. 15–49.
Moriarty, Jeffrey. 2003. "Against the Asymmetry of Desert." *Nous*, Vol. 37, No. 3, pp. 518–36.
Morris, Christopher W. 1991. "Punishment and Loss of Moral Standing." *Canadian Journal of Philosophy*, Vol. 21, No. 1, pp. 53–80.
Morris, Herbert. 1965. "Punishment for Thoughts." *The Monist*, Vol. 49, pp. 342–76.
1968. "Persons and Punishment." *The Monist*, Vol. 52, pp. 475–501.
1981. "A Paternalistic Theory of Punishment." In Murphy, ed. *Punishment and Rehabilitation*, pp. 42–54.
Mundle C. W. K. 1954. "Punishment and Desert." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 65–81.
1968. "Postscript." In Acton, ed., *Philosophy of Punishment*, 1969, pp. 81–2.

Murphy, Jeffrie G. 1973a. "Marxism and Retributivism." In Simmons et al., eds., *Punishment*, 1994, pp. 3–29.
(ed.). 1973b. *Punishment and Rehabilitation*, Belmont, CA: Wadsworth.
1985. "Retributivism, Moral Education and the Liberal State". In Murphy, *Retribution Reconsidered: More Essays in the Philosophy of Law*. Dordrecht: Kluwer Academic Publishers, 1992, pp. 15–30.
1987. "Does Kant Have a Theory of Punishment?" In Murphy, *Retribution Reconsidered*, 1992, pp. 31–60.
1988a. "The Retributive Emotions." In Murphy and Hampton. *Forgiveness and Mercy*, pp. 1–10.
1988b. "Hatred: A Qualified Defense." In Murphy and Hampton, *Forgiveness and Mercy*, pp. 88–110.
1990a. "Why Have Criminal Law at All?" in Murphy, *Retribution Reconsidered*, 1992, pp. 1–13.
1990b. "Getting Even: The Role of the Victim" in Murphy, *Retribution Reconsidered*, 1992, pp. 61–85.
1990c. "Review of George Sher, *Desert*." *Philosophical Review*, Vol. 99, No. 2 (April), pp. 280–3.
Murphy, Jeffrie and Jean Hampton. 1988. *Forgiveness and Mercy*. Cambridge: Cambridge University Press.
Narveson, Jan. 1974. "Three *Analysis* Retributivists." *Analysis*, Vol. 38, pp. 194–9.
Nathanson, Stephen. 1985. "Does It Matter If the Death Penalty Is Arbitrarily Administered?" *Philosophy and Public Affairs*, Vol. 14, pp. 149–64.
New, Christopher. 1992. "Time and Punishment." *Analysis*, Vol. 52, pp. 35–40.
1995. "Punishing Times: Reply to Smilansky." *Analysis*, Vol. 55, pp. 60–2.
Newman, Graeme. 1978. *The Punishment Response*. Philadelphia: J.B. Lippincott.
1983. *Just and Painful: A Case for the Corporal Punishment of Criminals*. London: Macmillan.
Nino, C. S. 1983. "A Consensual Theory of Punishment." In Simmons et al., eds., *Punishment*, 1994, pp. 94–111.
1986. "Does Consent Override Proportionality?" *Philosophy and Public Affairs*, Vol. 15, No. 2 (Spring), pp. 183–7.
1991. *The Ethics of Human Rights*. Oxford: Clarendon Press.
Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
Oldenquist, Andrew. 1986. "The Case for Revenge." *The Public Interest*, Vol. 82, pp. 72–80.
1988. "An Explanation of Retribution." *The Journal of Philosophy*, Vol. 85, No. 9 (September), pp. 464–78.
Olsaretti, Serena, ed. 2003. *Desert and Justice*. Oxford: Clarendon press.
Otsuka, Michael. 1996. "Quinn on Punishment and Using Persons as a Means." *Law and Philosophy*, Vol. 15, No. 2, pp. 201–8.
2003. *Libertarianism without Inequality*. Oxford: Clarendon Press.
Parfit, Derek. 1986. "Overpopulation and the Quality of Life." In Peter Singer, ed., *Applied Ethics*. Oxford: Oxford University Press, pp. 145–64.

Paton, Margaret. 1979. "Can an Action Be Its Own Punishment?" *Philosophy*, Vol. 54, pp. 534–40.
Pearson, Frank S. 1987. "Taking Quality into Account: Assessing the Benefits and Costs of New Jersey's Intensive Supervision Program." In McCarthy, ed., *Intermediate Punishments*, pp. 85–97.
Perkins, Lisa H. 1970. "Suggestion for a Justification of Punishment." *Ethics*, Vol. 81, No. 1, pp. 55–61.
Petersilia, Joan. 1987. "Georgia's Intensive Probation: Will the Model Work Elsewhere?" In McCarthy, ed., *Intermediate Punishments*, pp. 15–30.
Pettus, Katherine Irene. 2005. *Felony Disenfranchisement in America: Historical Origins, Institutional Racism, and Modern Consequences*. New York: LFB Scholarly Publishing LLC.
Philips, Michael. 1985. "The Inevitability of Punishing the Innocent." *Philosophical Studies*, Vol. 48, pp. 389–91.
1986. "The Justification of Punishment and the Justification of Political Authority." *Law and Philosophy*, Vol. 5, pp. 393–416.
Pilon, Roger. 1978. "Criminal Remedies: Restitution, Punishment, or Both?" Vol. 88, No. 4, pp. 348–57.
Pojman, Louis. 1999. "Merit: Why Do We Value It?" *Journal of Social Philosophy*, Vol. 30, No. 1, pp. 83–102.
Pratt J. R. 1968. "Professor Ducasse and the Meaning of Punishment." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 20–5.
Primorac [sic], Igor. 1978. "Utilitarianism and Self–Sacrifice of the Innocent." *Analysis*, Vol. 34, pp. 185–93.
1981. "Is Retributivism Analytic?" *Philosophy*, Vol. 56, pp. 203–11.
Primoratz, Igor. 1989a. *Justifying Legal Punishment*. Atlantic Highlands, NJ, and London: Humanities Press International.
1989b. "Punishment as Language." *Philosophy*, Vol. 64, pp. 187–205.
Prust, Richard C. 1988. "How to Treat a Criminal." *Public Affairs Quarterly*, Vol. 2, No. 3 (July), pp. 33–50.
Quinn, Warren. 1985. "The Right to Threaten and the Right to Punish." In Simmons et al., eds., *Punishment*, 1994, pp. 47–93.
Quinton, Anthony M. 1954. "On Punishment." In Acton, ed., *The Philosophy of Punishment*, 1969, pp. 55–64.
Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review*, Vol. 64, No. 1, pp. 3–32.
Regan, Peter F. 1968. "Discussion of 'Immorality, Crime and Treatment'." In Madden et al., eds., *Philosophical Perspectives on Punishment*, pp. 52–5.
Reiff, Mark R. 2005. *Punishment, Compensation, and Law: A Theory of Enforceability*. Cambridge: Cambridge University Press.
Reitan, Eric. 1996. "Punishment and Community: The Reintegrative Theory of Punishment." *Canadian Journal of Philosophy*, Vol. 26, No. 1, pp. 57–82.
"Restitution Sought in Guard's Stabbing." 2003. *New York Times* (National Edition), April 24, p. A29.
Rosen, F. 1997. "Utilitarianism and the Punishment of the Innocent: The Origins of a False Doctrine." *Utilitas*, Vol. 9, No. 1, pp. 23–37.

Rothbard, Murray. 1982. *The Ethics of Liberty*. Atlantic Highlands, NJ: Humanities Press.

Sadurski, Wojciech. 1985. *Giving Desert Its Due: Social Justice and Legal Theory*. Dordrecht: D. Reidel.

1989. "Theory of Punishment, Social Justice, and Liberal Neutrality." *Law and Philosophy*, Vol. 7, pp. 351–73.

Satre, Thomas W. 1987–8. "Contracting for Punishment." *Philosophy Research Archives*, Vol. 13, pp. 431–8.

Sayre-McCord, Geoffrey. 2001. "Criminal Justice and Legal Reparations as an Alternative to Punishment." *Philosophical Issues*, Vol. 11, pp. 502–29.

2002. "In Defense of Reparations: A Reply to Estlund and Gaus." In Enrique Villanueva, ed., *Legal and Political Philosophy, Social, Political and Legal Philosophy*, Vol. 1. (New York: Rodopi), pp. 371–83.

Schafer, Stephen. 1970. *Compensation and Restitution to Victims of Crime*, second edition. Montclair, NJ: Patterson Smith.

Schedler, George. 1976. "On Telishing the Guilty." *Ethics*, Vol. 86, No. 3, pp. 256–60.

Scheffler, Samuel. 2003. "Distributive Justice and Economic Desert." In Olsaretti, ed., *Desert and Justice*, pp. 69–91.

Scheid, Don E. 1990. "Note on Defining 'Punishment'." *Canadian Journal of Philosophy*, Vol. 10, No. 3, pp. 453–62.

1995a. "Davis, Unfair Advantage Theory, and Criminal Dessert." *Law and Philosophy*, Vol. 14, pp. 375–409.

1995b. "Review of Michael Davis, *To Make the Punishment Fit the Crime*." *Ethics*, Vol. 105, No. 3, pp. 667–70.

1997. "Constructing a Theory of Punishment, Desert, and the Distribution of Punishments." *Canadian Journal of Law and Jurisprudence*, Vol. 10, No. 2, pp. 441–506.

Schmidt, Annesley K. and Christine E. Curtis. 1987. "Electronic Monitors." In McCarthy, ed., *Intermediate Punishments*, pp. 137–52.

Sendor, Benjamin B. 1996. "The Relevance of Conduct and Character to Guilt and Punishment." *Notre Dame Journal of Law, Ethics and Public Policy*, Vol. 10, pp. 99–136.

Shafer-Landau, Russ. 1991. "Can Punishment Morally Educate?" *Law and Philosophy*, Vol. 10, pp. 189–219.

2000. "Retributivism and Desert." *Pacific Philosophical Quarterly*, Vol. 81, pp. 189–214.

Sher, George. 1987. *Desert*. Princeton, NJ: Princeton University Press.

1997. "Deserved Punishment Revisited." In Sher, *Approximate Justice: Studies in Non–Ideal Theory*. Lanham, MD: Rowman & Littlefield, pp. 165–80.

Shook, John. 2004. "Dewey's Rejection of Retributivism and His Moral-Education Theory of Punishment." *Journal of Social Philosophy*, Vol. 35, No. 1, pp. 66–78.

Simmons, A. John. 1991. "Locke and the Right to Punish." In Simmons et al., eds., *Punishment*, 1994, pp. 219–60.

Simmons, A. John, Marshall Cohen, Joshua Cohen, and Charles R. Beitz, eds. 1994. *Punishment*. Princeton, NJ: Princeton University Press.

Sink, Mindy. 2002. "New Mexico: Ignition Locks for D.W.I. Offenders." *New York Times* (National Edition), March 8, p. A15.
Skinner, B. F. 1953. *Science and Human Behavior*. New York: Macmillan.
Smart J. J. C. 1973. "An Outline of a System of Utilitarian Ethics." In Smart and Bernard Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, pp. 3–76.
Smilansky, Saul. 1990. "Utilitarianism and the 'Punishment' of the Innocent: The General Problem." *Analysis*, Vol. 50, No. 4, pp. 256–61.
1992. "Two Apparent Paradoxes about Justice and the Severity of Punishment." *The Southern Journal of Philosophy*, Vol. 30, No. 3, pp. 123–8.
1994. "The Time to Punish." *Analysis*, Vol. 54, pp. 50–3.
Snook, I. A. 1983. "Scheid on Punishment." *Canadian Journal of Philosophy*, Vol. 8, No. 1, pp. 129–32.
Sorrell, Tom. 1999. "Punishment in a Kantian Framework." In Matravers, ed., *Punishment and Political Theory*, pp. 10–27.
Statman, Daniel. 1997. "The Time to Punish and the Problem of Moral Luck." *Journal of Applied Philosophy*, Vol. 14, No. 2, pp. 129–35.
Stephenson, Wendell. 1990. "Fingarette and Johnson on Retributive Punishment." *The Journal of Value Inquiry*, Vol. 24, pp. 227–33.
Sterba, James P. 1977. "Retributive Justice." *Political Theory*, Vol, 5, No. 3, pp. 349–62.
1990. "A Rational Choice Theory of Punishment." *Philosophical Topics*, Vol. 18, No. 1, pp. 171–81.
Stern, Laurence. 1970. "Deserved Punishment, Deserved Harm, Deserved Blame." *Philosophy*, Vol. 45, pp. 317–27.
Strang, Heather. 2002. *Repair or Revenge: Victims and Restorative Justice*. Oxford: Clarendon Press.
Strickland, Ruth Ann. 2004. *Restorative Justice*. New York: Peter Lang.
Strong, Edward W. 1969. "Justification of Juridical Punishment." *Ethics*, Vol. 79, No. 3, pp. 187–98.
Sullivan, Dennis and Larry Tifft. 2001. *Restorative Justice: Healing the Foundations of Our Everyday Lives*. Monsey, NY: Willow Tree Press.
Sverdlik, Steven. 1988. "Punishment." *Law and Philosophy*, Vol. 7, pp. 179–201.
Teichman, Jenny. 1973. "Punishment and Remorse." *Philosophy*, Vol. 48, pp. 335–46.
Ten, C. L. 1987. *Crime, Guilt, and Punishment: A Philosophical Introduction*. Oxford: Clarendon Press.
1990. "Positive Retributivism." *Social Philosophy and Policy*, Vol. 7, No. 2 (Spring), pp. 194–208.
2000. "Deserved Punishment and Benefits to Victims," *Utilitas*, Vol. 12, No. 1, pp. 85–90.
Thomson, Judith Jarvis. 1980. "Rights and Compensation." *Nous*, Vol. 14, No. 1, pp. 3–15.
Tudor, Steven. 2001. "Accepting One's Punishment as Meaningful Suffering." *Law and Philosophy*, Vol. 20, pp. 581–604.
Tunick, Mark. 1992. *Punishment: Theory and Practice*. Berkeley: University of California Press.

Twentieth Century Fund Task Force on Criminal Sentencing. 1976. *Fair and Certain Punishment: Report of the Twentieth Century Fund Task Force on Criminal Sentencing*. New York: McGraw–Hill.

Vaughn, Joseph B. 1987. "Planning for Change: The Use of Electronic Monitoring as a Correctional Alternative." In McCarthy, ed., *Intermediate Punishments*, pp. 153–68.

Von Hirsch, Andrew. 1976. *Doing Justice: The Choice of Punishments* (Report of the Committee for the Study of Incarceration). New York: Hill and Wang.

1993. *Censure and Sanctions*. Oxford: Clarendon Press.

Von Hirsch, Andrew, Julian V. Roberts, and Anthony Bottoms, eds. 2003. *Restorative Justice and Criminal Justice: Competing or Reconcilable Paradigms?* Oxford and Portland, OR: Hart.

Walker, Mark Thomas. 1993. "Punishment – A Tale of Two Islands." *Ratio* (New Series), Vol, 6, No. 1 (June), pp. 63–71.

Walker, Nigel. 1993. *Why Punish?: Theories of Punishment Reassessed*. Oxford: Oxford University Press.

1995. "Nozick's Revenge." *Philosophy*, Vol. 70, pp. 581–6.

1999. "Even More Varieties of Retribution." *Philosophy*, Vol. 74, pp. 595–605.

Weiser, Benjamin. 2002. "Prime Time and Punishment: A Sentence of 10 Months without TV Is Blocked." *New York Times* (National Edition), March 7, p. A24.

Wertheimer, Alan. 1976. "Deterrence and Retribution." *Ethics*, Vol. 86, pp. 181–99.

White, G. Edward. 2003. *Tort Law in America*, expanded edition. Oxford: Oxford University Press.

Wilkinson, Stephen. 1996. "Restitution without Punishment: Is It Enough to Make Criminals Pay?" In Henry Tam, ed., *Punishment, Excuses and Moral Development*. Aldershot, UK: Avebury, pp. 35–53.

Wilson, John. 1983. "The Purposes of Retribution." *Philosophy*, Vol. 58, pp. 521–7.

Winch, Peter. 1972. "Ethical Reward and Punishment." In Winch, *Ethics and Actions*. London: Routledge & Kegan Paul, pp. 210–28.

Wolgast, Elizabeth H. 1985. "Intolerable Wrong and Punishment." *Philosophy*, Vol. 60, pp. 161–74.

Wood, David. 1997. "Reductivism, Retributivism, and the Civil Detention of Dangerous Offenders." *Utilitas*, Vol. 9, No. 1 (March), pp. 131–46.

Wright, Martin. 1996. *Justice for Victims and Offenders: A Restorative Response to Crime*, second edition. Winchester, UK: Waterside Press

Zehr, Howard. 2002. *The Little Book of Restorative Justice*. Intercourse, PA: Good Books.

Zimmerman, Michael J. 1995. "Review of Peter A. French, *Responsibility Matters*, Jeffrie G. Murphy, *Retribution Reconsidered: More Essays in the Philosophy of Law*, and George Sher, *Desert.*" *Nous*, Vol. 29, No. 2, pp. 248–59.

Index

Abel, Charles F., 27n33, 215n4, 216n7, 217n9, 231n19, 241n28, 250, 259–60n39
act versus outcome objection, 101–3
Adler, Jacob, 6n4, 7, 11, 20n27, 119n32
Adler, Matthew D., 184n22
Alexander, Larry, 94n11, 168, 202
Anas, Brittany, 53
Anderson, Jami L., 135n51
Applebaum, Anne, 14n16
Aquinas, Thomas, 6n4
Aristotle, 6n4
Armstrong, K. G., 41n4, 44n9, 54n20, 61n25, 188n28
authorization requirement, 23–4, 40
Avio, K. L., 82

Bagaric, Mirko, 17n20
Baier, Kurt, 18n23
Baker, Brenda M., 172n10
Barnett, Randy, 2, 27, 28, 28n35, 31n42, 102n18, 216, 217n9, 220n10, 224–6, 228, 231, 246, 259n39, 262n43, 264
Barton, Charles K. B., 33n46, 152
Baylis, Charles A., 27n33, 214n2
Bean, Philip, 14n14, 100n16, 214
Beccaria, Cesare, 38
Bedeau, Hugo Adam, 31n41
Bell v. Wolfish, 15n18
Bello, Edward, 180–1
Benn, Stanley I., 6n4, 13, 23, 43, 50n17, 63n27, 233n24
Bennett, Lawrence A., 232n21
Bentham, Jeremy, 38, 43n7, 47n13, 59
Berenson, Alex, 162n5
Bickenbach, Jerome E., 172n10
Bird, Larry, 10n9
Blanshard, Brand, 93n10, 98n14
Blomberg, Thomas G., 232n20
Blume, Delorys, 214n2
Blume, Robert, 214n2
Blumoff, Theodore, 29n36, 208n40
Bok, Sissela, 144
Bottoms, Anthony, 217n8
Bradley, Gerard V., 119n32, 124n37
Brady, James B., 106n25
Braithwaite, John, 40n2, 41n4, 54n20, 83–4n36, 110n26, 123n35, 127n41, 140n57, 168n9, 184n23, 208n40
Brandt, R. B., 44n9, 53, 59n24, 60, 63n27, 76n31
Brunk, Conrad G., 82n34
Burcroff, Lisa C., 232n20
Burgh, Richard W., 14n14, 110n26, 125n38, 168n9

Carcasole, Joseph, 198n36, 202n39
Cederblom, Jerry, 123n36, 196–7
Champlin, T. S., 43n7
Charvet, John, 22n28

Clark, Michael, 14n14
Clear, Todd R., 232n21
consent solution, 105, 156–71, 211
consequentialism, *see* nonutilitarian consequentialism; utilitarianism
Corlett, J. Angelo, 6n4, 14n14, 104n21
Cottingham John, 86n3
Cragg, Wesley, 41n4, 54n20, 82n34
Curtis, Christine E., 232n20, 233, 234

Dagger, Richard, 27n33, 119, 120, 124–6, 217, 228, 235, 245n31, 249, 250, 255, 268n47
Davis, Michael, 23n29, 86n3, 129–34, 142n58
Day, J. P., 120n32
Deigh, John, 181n18
del Vecchio, Giorgio, 216
Dennett, Daniel, 47n13
Denning, Lord, 171n10
desert, *see* retributivism, desert-based
Dewey, John, 181n18
Dimock, Susan, 18n23, 51n18, 102n18, 123n35, 131n49, 144n60, 145–9
disproportionate punishment objection
  as objection to act-utilitarian solution, 54–8
  as objection to consent solution, 162–4, 168–70
  as objection to forfeiture-based retributivist solution, 112–14
  as objection to motive-utilitarian solution, 78
  as objection to nonutilitarian consequentialist solutions, 81–3
  as objection to rule-utilitarian solution, 76–7
  as objection to self-defense solution, 201–3
  as problem for passing foundational test, 35
Dolinko, David, 100n16, 103n20, 126n39, 129n46, 131n49, 132n50, 188n28, 191n30
Dostoevsky, Fyodor, 89
Doyle, James F., 119–20n32, 208n40
Ducasse, C. J., 14n14, 19n24
Duff, R. A., 11n11, 22, 33n45, 70n30, 123, 132n50, 172–4, 181n18, 183n21, 186

Ellin, Joseph, 216n7, 217n9, 231–2n19, 251n36, 264, 267n44, 270n49
Ellis, Anthony, 86n3, 102n18, 131n49, 139n55, 194n32
entailment test, 34, 35–6
Ewing, A. C., 18n23, 181n18
Ezorsky, Gertrude, 123n35

fairness, *see* retributivism, fairness-based
Farrell, Daniel M., 196, 197n34, 202
Fatic, Aleksandar, 14n14, 143n59, 215, 217n8, 259n38
Feinberg, Joel, 22n28, 172n10, 222n13
Ferraro, Geraldine, 9n8
Fingarette, Herbert, 30–1, 33n45, 135n51
Finnis, John, 119, 121, 122
Flew, Antony, 6n4, 18n21
Flynn, Suzanne, 232n21
forfeiture claim, 105–7
  see also retributivism, forfeiture-based
Fouche, Fidela, 135n51
foundational test, 34–5
free rider
  nonoffender as, 140–1
  offender as, 120–2
  offender as not, 122–35
  rights of, 138–9

Gaffney, Paul, 31n37
Gahringer, Robert E., 181n18, 183, 190
Garcia, J. L. A., 89n6, 137n53
Garland, D., 22
Gavison, Ruth, 41n4, 53n19, 54n20, 143n59

Gendin, Sidney, 18n23, 92n9, 152
Gerstein, Robert S., 119n32, 152
Gert, Bernard, 127n41
Gert, Heather J., 14n14
Gewirth, Alan, 119n32
Golash, Deirdre, 14n14, 40n2, 41n4, 86n2, 102n18, 184n22, 190n29, 194n32, 217n9, 231n19, 234n25, 268n46
Goldman, Alan H., 105–9, 111, 168n9

Hajdin, Mane, 216n7, 267n44
Haksar, Vinit, 105, 106n25
Hampton, Jean, 11n11, 54n20, 93n10, 123n35, 125n38, 172n10, 173n12, 177, 181–92
Hand, Michael, 14n14
Hare, R. M., 18n23, 45n11, 63n27
harm requirement, 6–12
harm versus punishment objection
  as objection to consent solution, 170–1
  as objection to debt-based retributivist solution, 151–2
  as objection to fairness-based retributivist solution, 141–3
  as objection to forfeiture-based retributivist solution, 117–19
  as objection to self-defense solution, 205–7
  as objection to trust-based retributivist solution, 147
Hart, H. L. A., 6n4, 20n27, 22n28, 43n7, 171, 208n40
Hawkins, D. J. B., 41n4, 54n20, 87n4
Hawkins, Gordon, 22n28
Hegel, Georg, 6n4, 119, 181n18
Hellerstein, Alvin K., 180–1
Hershenov, David B., 123n35, 184, 233n23, 261n42, 270n48
Hill, Thomas E. Jr., 22n28, 85, 86n1, 93n10, 102n18, 214
Hobbes, Thomas, 6n4, 105
Hobson, Peter, 184n24
Hoekema, David, 27n33, 144–5, 148, 186n27, 225, 235n26, 259n38, 261n42
Holland, Gina, 15n18
Holmgren, Margaret R., 233, 270n48
Honderich, Ted, 6n4
Hooker, Brad, 63n27
Hume, David, 105
Hurka, Thomas, 194n32
Husak, Douglas N., 22n28, 184n22
hybrid solutions, 207–12
  conjunctive versions, 208–11
  disjunctive versions, 211–12

intending harm requirement, 12–17, 27n34, 35, 40, 118, 148, 151–2, 170–1, 179, 185n26, 206, 233–4

Johnson, Oliver A., 31n43
justice
  distributive, 120–2, 196–7, 200–1, 204–7
  retributive, *see* retributivism

Kant, Immanuel, 6n4, 86n1, 119, 120n32, 214
Kasachkoff, Tziporah, 8n7, 14n14, 18n23
Kemp, Kenneth W., 17n20
Kennedy, Anthony, 15n18
Kershnar, Stephen, 87–93, 96n12, 99, 101, 105, 135n51, 137n53, 161n4
Khatchadourian, Haig, 181n18
Kidder, Joel, 208n40
Kionka, Edward J., 220n10, 221n11, 223n14, 231n18
Kleinberg, Stanley S., 217, 225n15
Kleinig, John, 16n19, 44n9, 87n4, 88, 89n7
Klimchuk, Dennis, 273n51
Knowles, Dudley, 63n27, 168n9
Koehl, Richard A., 214n2, 259n38
Korman, Dan, 143n59, 145n61, 146n63, 147n65

Lacey, Nicola, 20n26, 100n16, 102n18
Latessa, Edward J., 232n21

Lemos, Ramon M., 181n18
Lewis, C. S., 41n4, 61n25, 87n4
Lippke, Richard L., 83n36, 110n26, 112, 113n28, 115n30, 122n34
Locke, Don, 18n23
Locke, John, 6n4, 105
Loewy, Arnold H., 27n33
logical entailment argument, 30–3, 89n7, 177, 190, 214
Long, Roderick T., 216n7, 243n29
Luke, Andrew, 215n3
Lyons, William, 44–5

Mabbott, J. D., 41n4, 188n28
Mackie, J. L., 89n6
Manser, A. R., 101n17
Marsh, Frank H., 27n33, 215n4, 216n7, 217n9, 231n19, 241n28, 250, 259–60n39
Marshall, James D., 184n24
Martin, Rex, 27n33, 29–30
Mason, Elinor, 79n33
Matravers, Matt, 84n36, 120n32, 208n40
McBain, Ed, 236n27
McCarthy, Belinda R., 233
McCloskey, H. J., 6n4, 24n31, 41n4, 48n14, 50n17, 214n2
McDermott, Daniel., 102n18, 106n25, 119n32, 149–52
McKerlie, Dennis, 194n32
McTaggart, J. Ellis, 181n18
Menninger, Karl, 214n2
Metz, Thaddeus, 172n10
Mill, John Stuart, 191
Miller, David, 22n28
Miller, Franklin G., 217, 225n15, 245n31, 248n34, 249, 255, 268n43
Miller, Vanessa, 13n13
Miller, W. A., 172n10
Montague, Phillip, 4, 196, 197n34, 198, 206, 216n5, 222n13
Moore, David B., 217n8
Moore, Michael S., 85, 87–93, 99–101, 103
moral education solution, 173n12, 180–92
moral luck, 229–31, 254–5
Moriarty, Jeffrey, 103n20
Morris, Christopher W., 105n22, 107n25
Morris, Herbert, 30n37, 119, 141–2n58, 181n18, 192n31
Mundle, C. W. K., 87n4, 89n6, 92
Murphy, Jeffrie, 61n25, 88, 119, 120n32, 122, 127n41, 142n58, 153, 190n29

Narveson, Jan, 93n10, 123n35
Nathanson, Stephen, 31n37
New, Christopher, 41n3
Newman, Graeme, 6n4, 16n19, 31n37
Nickel, Jim, 19n25
Nino, C. S., 14n14, 22n28, 55, 61n25, 105, 157–9, 161–2, 164–5, 168–9
no excuses objection
  as objection to act-utilitarian solution, 58–60
  as objection to fairness-based retributivist solution, 137–8
  as objection to motive-utilitarian solution, 78
  as objection to rule-utilitarian solution, 76–7
  as objection to self-defense solution, 203–5
  as problem for passing foundational test, 59
nonutilitarian consequentialism, 79–84, 153–4, 182n19, 197–8n33
not punishing the guilty objection
  as objection to act-utilitarian solution, 52–4
  as objection to conjunctive hybrid solutions, 209
  as objection to consent solution, 160–5
  as objection to debt-based retributivist solution, 151
  as objection to desert-based retributivist solution, 93–8
  as objection to disjunctive hybrid

solutions, 211
as objection to fairness-based retributivist solution, 122–37
as objection to forfeiture-based retributivist solution, 115n31
as objection to moral education solution, 184–8
as objection to nonutilitarian consequentialist solutions, 81–3
as objection to reprobative solution, 176–80
as objection to rule-utilitarian solution, 75–6
as objection to trust-based retributivist solution, 147
as problem for passing entailment test, 53–4
Nozick, Robert, 92n9

Oldenquist, Andrew, 33n46, 173n11
Otsuka, Michael, 197–8n33

Parfit, Derek, 49n16
paternalism, 190–2, 245–7
Paton, Margaret, 24n31
Pearson, Frank S., 232n21
Perkins, Lisa, 82, 154
Peters, Richard S., 63n27
Petersilia, Joan, 232n21
Pettit, Philip, 40n2, 41n4, 54n20, 83–4n36, 110n26, 123n35, 127n41, 140n57, 168n9, 184n23,
Pettus, Katherine Irene, 15n17
Philips, Michael, 29n36, 49n15
Pilon, Roger, 105, 106n25, 217, 259n38
Plato, 6n4, 181n18
Pojman, Louis, 92n9
Pratt, J. R., 18n23
Primoratz, Igor, 6n4, 14n14, 20n27, 23, 31n37, 42n6, 47n13, 53n19, 54n20, 58n22, 59n23, 70n30, 85, 87n4, 88, 172n10, 173n11, 177, 182–5, 188n28
principle of separability argument, 29–30
probation, 11n10
Prust, Richard C., 181n18
punishing the innocent objection
as objection to act-utilitarian solution, 41–52
as objection to consent solution, 165–8
as objection to desert-based retributivist solution, 99–101
as objection to disjunctive hybrid solutions, 211
as objection to fairness-based retributivist solution, 140–1
as objection to forfeiture-based retributivist solution, 115–17
as objection to moral education solution, 188–90
as objection to motive-utilitarian solution, 78
as objection to nonutilitarian consequentialist solutions, 81–3
as objection to reprobative solution, 180
as objection to revenge-based retributivist solution, 273n51
as objection to rule-utilitarian solution, 64–75
as objection to self-defense solution, 198–201
as objection to trust-based retributivist solution, 144–6
as problem for passing entailment test, 50–2
as problem for passing foundational test, 48–50
possibility of, 18–19, 43
vicarious version of, 45–6, 67–8, 78, 81, 198–201
punishment
community service as form of, 11–12
definition of, 3–26
of children by parents, 7–8, 14, 184, 190
problem of, 1, 28–33, 66, 75, 116, 118–19, 159, 213, 215
versus restitution, 26–8, 215, 220, 229, 233–4
versus reward, 9, 12, 18–19
versus self-defense, 193–4
versus telishment, 65

Quinn, Warren, 195–200, 202, 206
Quinton, Anthony M., 18n21, 19, 20n26

Radzik, Linda, 14n14
Rawls, John, 63–77
Regan, Peter F., 214n2
Reiff, Mark R., 24n31, 30n37, 162, 163
Reitan Eric, 208n40
reprobative requirement, 21–3, 40, 171
reprobative solution, 171
restitution
  pure versus punitive, 27–8
  theory of pure, 215–75
    see also restorative justice; punishment versus restitution
restorative justice, 216–17n8
    see also restitution
retributive requirement, 17–21, 40, 51
retributivism
  debt-based, 149–52, 209
  desert-based, 87–103, 152n67, 180, 185n25, 211
  fairness-based, 119–43
  forfeiture-based, 103–19, 160n2, 165n7, 211
  revenge-based, 152–4, 269–74
  trust-based, 143–9
Roberts, Julian V., 217n8
Rosen, F., 43n7, 63n27
Rothbard, Murray, 105
Royko, Mike, 88

Sadurski, Wojciech, 119n32, 121n33, 127n40, 127–8n42, 135–6n52, 138n54
Satre, Thomas W., 102n18
Sayre-McCord, Geoffrey, 217n8
Schafer, Stephen, 215n4
Schedler, George, 18n21
Scheffler, Samuel, 177n15
Scheid, Don E., 14n14, 18n23, 131n49
Schmidt, Annesley K., 232n20, 233, 234
self-defense solution, 192–207
Sendor, Benjamin B., 96n12
Shafer-Landau, Russ, 93n10, 123n35, 182n20
Shapiro, Carol, 232n21
Sher, George, 119, 127–9, 130, 134, 135–7, 141–2n58
Sherwin, Emily, 94n11
Shook, John, 181n18
Simmons, A. John, 105n22, 106n25, 115n30
Sink, Mindy, 233n22
Skinner, B. F., 214n2
Smart, J. C. C., 38, 47
Smilansky, Saul, 41n3, 442n5, 55n21
Snook, I. A., 8n7
Sorrell, Tom, 120n32
Statman, Daniel, 41n3
Stephen, James Fitzjames, 153–4
Stephenson, Wendell, 14n14, 31n43
Sterba, James P., 120n32
Stern, Laurence, 100n16
Stewart, Martha, 162n5
Strang, Heather, 215n4, 217n8, 268
Strickland, Ruth Ann, 215n4, 217n8
Strong, Edward W., 208n40
Sullivan, Dennis, 217n8
Sverdlik, Steven, 14n14, 17n20, 48n14

Teichman, Jenny, 213n1
telishment, 65–9
Ten, C. L., 31n39, 43, 44n9, 58n22, 59, 70n30, 128n43, 129n46, 135n51, 184, 186n27
theory of pure restitution, 215–75
  historical practices and, 31n42
  pure versus punitive restitution, 27–8
Thomson, Judith Jarvis, 222n13
Tifft, Larry, 217n8
Torpy, Eric James, 10n9
treating people as a means
  and moral education solution, 187
  as objection to act-utilitarian solution, 60–2

as objection to motive-utilitarian solution, 79
as objection to nonutilitarian consequentialist solutions, 81–3
as objection to rule-utilitarian solution, 77
Tudor, Steven, 172n10
Tunick, Mark, 9n8, 15n18, 173n11, 217, 225n15, 235n26, 245, 248n34, 259n38

utilitarianism, 39–79
act-version, 39–62, 209, 211
motive-version, 77–9
rule-version, 62–77, 208n41

Von Hirsch, Andrew, 22n28, 85, 86n3, 119n32, 123n35, 172n10, 173n11, 178, 186n27, 217n8
Vaughn, Joseph B., 232n20

Waldo, Gordon P., 232n20
Walker, Mark Thomas, 235
Walker, Nigel, 16n19, 61–2, 86n3, 92n9, 123n35
Weiser, Benjamin, 181
Wertheimer, Alan, 63n27
White, G. Edward, 221n11, 231n18
Wilkins, Leslie T., 85
Wilkinson, Stephen, 216n7, 229, 255, 260n40
Wilson, John, 216n7
Winch, Peter, 24n31
Wolgast, Elizabeth H., 102n18
Wood, David, 233n24
Wright, Martin, 14n14, 31n42, 216n7, 268n47

Zehr, Howard, 217n8
Zimmerman, Michael J., 127n41, 128n44