

Washington University in St. Louis

Washington University Open Scholarship

All Theses and Dissertations (ETDs)

January 2011

The Moral Permissibility of Punishment

Zachary Hoskins

Washington University in St. Louis

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>

Recommended Citation

Hoskins, Zachary, "The Moral Permissibility of Punishment" (2011). *All Theses and Dissertations (ETDs)*. 156.

<https://openscholarship.wustl.edu/etd/156>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Philosophy

Dissertation Examination Committee:

Larry May (co-chair)

Christopher Heath Wellman (co-chair)

Anne Margaret Baxley

Marilyn Friedman

Frank Lovett

Ian MacMullen

THE MORAL PERMISSIBILITY OF PUNISHMENT

by

Zachary Wallace Hoskins

A dissertation presented
to the Graduate School of Arts and Sciences
of Washington University in St. Louis
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

May 2011

St. Louis, Missouri

Copyright by
Zachary Wallace Hoskins
2011

ABSTRACT

This dissertation offers an account of the moral permissibility of criminal punishment. Punishment presents a distinctive moral challenge in that it involves a community's inflicting harm on individuals, treating them in ways that would typically be morally wrong. We can distinguish a number of different questions of punishment's permissibility. This dissertation focuses on four central questions: (1) Why may we punish? Why is it in principle permissible to inflict harm on criminal offenders? (2) Why should we punish? Is there a compelling reason to do so? (3) How may we punish? What principles should constrain impositions of punishment? And finally, (4) who is properly subject to punishment? Rather than expect to answer all of these questions by appeal to the same moral principle, this dissertation contends that the questions should be seen as distinct, and thus as appropriately answered by appeal to distinct moral considerations. Ultimately, the dissertation concludes that an institution of punishment that aims at deterrence, constrained by considerations of retribution and reform, is permissible insofar as the institution is among the mutually beneficial practices with which community members have reciprocal, fairness-based obligations to comply.

ACKNOWLEDGMENTS

I am especially grateful, first, to my dissertation advisors, Larry May and Kit Wellman, for their unflagging guidance and support. Both were extremely generous with their time, and their distinct styles as philosophers and advisors complemented each other well and proved a great benefit to me. In addition, I thank the other four members of my dissertation committee: Anne Margaret Baxley, Marilyn Friedman, Frank Lovett, and Ian MacMullen. All of my committee members offered not only valuable feedback during my dissertation defense, but also encouragement and friendship along the way.

A number of people provided helpful written comments on various chapters of the dissertation: Nate Adams, Anne Margaret Baxley, Richard Dagger, Antony Duff, Marilyn Friedman, Jason Gardner, Larry May, Victor Tadros, Krista Thomason, Carl Wellman, Kit Wellman, and David Wood. In addition, I appreciate the feedback I received from audience members at the 2008 Midsouth Philosophy Conference; the 2008, 2009, and 2010 International Social Philosophy Conferences; the 2010 APA Eastern Division Meetings; and two Washington University Political Theory Workshops, in 2008 and 2010.

The Washington University Graduate School of Arts and Sciences provided generous funding during my graduate studies, in particular a dissertation fellowship during my fifth year that allowed me to pursue my research without any teaching assistant responsibilities. I also appreciate the Graduate School and the Department of Philosophy for funding support as I traveled to conferences to present my work.

Chapter 1 was originally published as “Fair Play, Political, Obligation, and Punishment,” *Criminal Law and Philosophy* 5:1 (January 2011): 53-71. Copyright © 2011 Springer Science+Business Media B.V. The final publication is available at www.springerlink.com. Chapter 2, “Deterrent Punishment and Respect for Persons,” was originally published in *Ohio State Journal of Criminal Law* 8:2 (Spring 2011).

Finally, and most of all, I thank my wife, Nora Wikoff, whose own research interest in criminal recidivism has dovetailed serendipitously with my work on punishment. Our discussions — over dinner, or while walking our dogs — have been invaluable in helping me clarify my own views about what sort of treatment we owe to criminal offenders. More importantly, I thank Nora for her unwavering patience, support, and generosity. This dissertation is dedicated to her.

Table of Contents

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
INTRODUCTION, Disaggregating the problem of punishment	1
I. Why may we punish?	3
II. Why should we punish?	5
III. How may we punish?	6
IV. Whom may we punish?	8
CHAPTER 1, Fair play and the in-principle permissibility of punishment	10
I. The standard fair play account and its drawbacks	13
II. An alternative fair play account	22
III. Objections	33
IV. Conclusion	45
CHAPTER 2, Deterrent punishment and respect for persons	47
I. Does deterrence use offenders as mere means?	50
II. Duff's critique of deterrence	59
III. Is deterrent punishment exclusionary?	62
IV. Does deterrent punishment offer the wrong sort of reasons for compliance?	67
V. Conclusion	72
CHAPTER 3, Retributivism as a constraint on punishment	73
I. The retributivist constraint, and the standard objection	76
II. Two ways the retributivist constraint may still be useful	87
III. Retributivism and respecting offenders as moral persons	94
IV. Conclusion	101
CHAPTER 4, Punishment, contempt, and the prospect of moral reform	103
I. What's wrong with contempt?	105
II. Treating offenders with respect	115
III. Reform to be promoted or not undermined?	122
IV. Objections	130
V. Conclusion	134

CHAPTER 5, Collective punishment and distribution of harms	136
I. The case for collective punishment	139
II. Distributing collective punishment's harms	142
III. Overriding net benefits of collective punishment?	157
IV. Nondistributive collective punishment	162
V. Conclusion	165
WORKS CITED	167

INTRODUCTION

Disaggregating the problem of punishment

This dissertation offers an account of the moral permissibility of the legal institution of punishment. As an institution, punishment presents a distinctive moral challenge in that it involves a political community's inflicting harm on some of its members, treating them in ways that would typically be morally wrong. Why are those who break the law subject to the state's imposing on them the sort of harms that are characteristic of punishment? Most theorists of punishment, like most members of the public, believe that punishment is indeed permissible. Nevertheless, providing a satisfactory explanation of why this is so has proven to be a thorny task, and this has led a number of theorists to argue in favor of abolishing the practice. Given how pervasive punishment has become — roughly one in every 31 American adults was in jail or prison, on probation, or on parole in 2008¹ — the prospect that the abolitionists may be right, that the practice may be morally impermissible, is a particularly troubling one. In this dissertation, however, I contend that punishment, properly constrained, can be permissible.

The question of punishment's moral permissibility is actually not one question but several. This dissertation focuses on four central questions: (1) Why may we punish? Why is it in principle permissible to inflict harm on criminal offenders? (2) Why should

¹ U.S. Department of Justice, "Total Correction Population," Bureau of Justice Statistics. Online at <http://bjs.ojp.usdoj.gov/index.cfm?ty=tp&tid=11> (accessed Oct. 8, 2010).

we punish? Is there a compelling reason to do so? (3) How (especially, how severely) may we punish? What principles should constrain impositions of punishment? And (4) who is properly subject to punishment?² Disaggregating several distinct questions of punishment is a strategy endorsed perhaps most notably by H. L. A. Hart, who distinguished the question of punishment's general justifying aim (i.e., why should we punish) from questions of its just distribution (i.e., who may be punished and how severely). Hart saw the seemingly intractable debate between consequentialists and retributivists about punishment's justification as a product of the tendency to oversimplify, to treat punishment as if it could be justified by some single moral principle or consideration. He wrote:

To counter this drive, what is most needed is *not* the simple admission that instead of a single value or aim (Deterrence, Retribution, Reform or any other) a plurality of different values and aims should be given as a conjunctive answer to some single question concerning the justification of punishment. What is needed is the realization that different principles (each of which may in a sense be called a 'justification') are relevant at different points in any morally acceptable account of punishment.³

Like Hart, I believe a defense of punishment requires answers to several distinct questions, answers that may appeal to different moral considerations.

² We might also ask who may permissibly do the punishing. I do not address this question explicitly, although it implicitly falls out of my account (in chapter 1) of punishment's in-principle permissibility that only the community's legal authority (rather than, say, the victims of the crime) may permissibly impose punishment.

³ H. L. A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford, U.K.: Oxford University Press, 1968), p. 3.

I. Why may we punish?

Although I agree with Hart about the importance of disaggregating various questions of punishment, I structure the questions somewhat differently than he did. Hart indicated that punishment's general justifying aim must be determined first, and that once this aim is known, it is then left to decide (based on distinct considerations) who may properly be punished, and how severely. As he put it, "in relation to any social institution, after stating what general aim or value its maintenance fosters we should enquire whether there are any and if so what principles limiting the unqualified pursuit of that aim or value."⁴ As a number of scholars have argued, however, a normatively prior question to punishment's aim is whether the practice itself is morally permissible. Given the value that liberal political communities such as ours place on individual autonomy, there appears to be a strong *prima facie* case against any institution that is centrally concerned with restricting individuals' freedoms. Defenders of punishment must explain, then, why a community's political authority can be justified in "treating some of its citizens in ways that it would be clearly wrong to treat others."⁵

Appeal to various aims will by itself typically be insufficient to justify punishment. We might think, for instance, that deterrence is a significant social benefit of punishment, but most of us would object to a practice that called for occasionally punishing innocent people even if this could be demonstrated to increase the deterrent impact significantly. Thus deterrence alone is an insufficient justification of the practice.

⁴ *Ibid.*, p. 10.

⁵ David Boonin, *The Problem of Punishment* (New York City: Cambridge University Press, 2008), p. 29.

What we first need is an account of why, in principle, the practice of imposing harms on people in the ways characteristic of punishment is permissible. If it is not, then appeal to whatever positive aims will be insufficient to justify the practice.

In chapter 1, I defend the claim that punishment is in principle permissible. Specifically, I develop a version of the fair play view, according to which the permissibility of punishment derives from reciprocal obligations shared by members of a political community. Because community members benefit from general compliance with the rules of the community, they incur a presumptive, fairness-based obligation to reciprocate by complying. On the standard fair play account, criminals gain an unfair advantage over other community members, and punishment is thus permissible as a means of removing this advantage. I contend, however, that this standard account is unsatisfying, largely because there is no advantage that an offender unfairly gains, in proportion to the seriousness of her crime, relative to other community members generally. Instead, I offer a more straightforward fair play account, according to which the rule instituting punishment as the response to crimes is itself among the community's mutually beneficial rules; as such, the rule instituting punishment is among those with which community members are presumptively obliged to comply. For offenders, compliance entails accepting being subject to punishment. This appeal to the reciprocal obligations that emerge in a political community explains why restricting the liberties of offenders is in principle permissible.

II. Why should we punish?

Even if punishment is in principle permissible, however, a full defense of the practice also requires an answer to the question of why we should want to punish. If there is no compelling reason to impose harms on certain individuals, no good that would come of it, then even if this institution can be shown to be permissible in principle, it may in practice be unjustified. Jeremy Bentham argued that “all punishment is mischief: all punishment in itself is evil. Upon the principle of utility, if it ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil.”⁶ Although I don’t share Bentham’s view that utilitarianism can ground a complete justification of punishment, I agree with him that an adequate defense of punishment requires that we provide a compelling answer to the question of why we should want to punish, of what good will come of it.

The fair play account I develop in chapter 1 depends on the claim that community members receive significant benefits from the institution of punishment. On my account, the chief benefit of such an institution is that it helps protect the security of community members by acting as a deterrent of crime. The aim of deterrence has frequently been criticized, however, as inconsistent with treating offenders with appropriate respect as persons. Opponents have leveled the broadly Kantian charge that deterrent systems of punishment use offenders as mere means to the social end of crime reduction, rather than respecting them as ends in themselves. This objection has been fleshed out in different

⁶ Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns and H. L. A. Hart (Oxford: Clarendon Press, 1996), p. 158.

ways, and in chapter 2 I consider three prominent versions of the objection: (1) that deterrent systems of punishment treat offenders as mere means to some social good (i.e., crime reduction); (2) that deterrent punishment implicitly excludes offenders from membership in the moral community; and (3) that such punishment offers community members the wrong sort of reasons for compliance with the law. I contend that, contrary to these various challenges, deterrence as an aim of punishment is in fact consistent with respecting offenders as moral persons.

III. How may we punish?

Next, even if the institution of punishment is in principle permissible, and even if we have some compelling and permissible reason(s) to want such an institution, particular impositions of punishment may still fail be impermissible if these punishments are excessive, or inhumane, or otherwise inappropriate. A defense of punishment thus requires an account of how punishments should be constrained, in their mode and severity, so as to treat offenders with respect as moral persons. Philosophers and legal theorists typically cite the retributivist principle that punishment should be only to the degree that is morally deserved as though this constraint is sufficient. The notion of desert, however, has proven notoriously difficult to flesh out. This has led critics to conclude that the retributivist constraint is of no use in answering the question of how we may punish. In chapter 3, I offer a partial defense of the retributivist constraint, arguing that the notion of desert may provide some, albeit imperfect, guidance in determining how to punish. Nevertheless, I contend that the retributivist constraint is insufficient to

ensure that offenders are treated with appropriate respect as moral persons. This is because retributivism focuses entirely on what is the morally deserved response to the given crime. In fact, however, what treatment a person deserves may also be a matter not only of what she has done, but also of who she is, and even of who she can be. Because retributivism evaluates desert by focusing only on the crime committed, in many cases (in particular, cases of the most serious crimes) the retributivist notion of desert may indicate punishments that many of us will regard as overly harsh. This indicates that the retributivist constraint is insufficient, and that some additional constraint is needed to ensure that punishments treat offenders with appropriate respect.

In chapter 4, I argue for such an additional constraint, one grounded in considerations of reform. I first flesh out a Kantian conception of contempt and highlight certain troubling features of contemptuous treatment. In particular, contempt is person- rather than act-focused; it is pervasive; it presents its object as inferior, if not altogether worthless; and it is cold and dismissive, i.e., it gives up on its object. Next I contend that punishments treat offenders with contempt if, in their mode or degree, they tend to undermine offenders' prospects for moral reform. On my account, such punishments are therefore impermissible. Unlike certain reform-based accounts, however, my view does not require that reform must be a positive aim of punishment — only that punishment should not tend to undermine the prospect of reform. Taken together, chapters 3 and 4 conclude that punishment should be constrained not only by retributivist considerations of desert, but also by considerations of moral reform.

IV. Whom may we punish?

Finally, we might ask who is properly subject to punishment. Generally, an answer to the in-principle permissibility question will imply an answer to this question of who may be punished. For instance, for those who contend that punishment is permissible because criminal wrongdoers forfeit their right against punishment, it will follow straightforwardly that punishment is permissible only of criminals (because only they have forfeited this right). Similarly, the fair play account I develop in chapter 1 implies that only criminal offenders are subject to punishment — they made themselves liable to punishment when they failed to comply with the mutually beneficial rules of the political community.

Even if we accept the fairly uncontroversial view that only criminal offenders are subject to punishment, however, a further question of who may be punished arises when we move from the context of domestic crimes to international crimes, such as crimes against humanity and genocide. Such crimes are perpetrated by groups; that is, they are made possible by the contributions of many individuals acting, to some extent, in concert. Some scholars have thus argued that punishments for international crimes should target the groups (typically states), as groups, rather than only targeting individual group members. There is a presumptive case against such a scheme, however, because of the very real danger that the *harms* of such punishments will distribute to all group members, many of whom may have played no role in (or even worked against) the crime. Collective punishment's advocates appear to have available three lines of response to this challenge: They may (1) argue that the harms of collective punishment can in fact be distributed

among group members in a justified way, (2) acknowledge that collective punishment's harms will distribute unjustly but contend that this presumptive injustice is overridden by the good that will be accomplished (or harm averted) by punishing groups as groups, or (3) contend that collective punishment can be imposed in ways such that the harms don't distribute among group members. In chapter 5, I examine each of these responses. I argue that none succeeds in overcoming the presumptive case against collective punishment.

CHAPTER 1

Fair play and the in-principle permissibility of punishment

Since H. L. A. Hart famously distinguished three different questions of criminal punishment — why should we punish, whom may we punish, and how much may we punish — responses to this disaggregation strategy have been mixed. Some have argued that it is *ad hoc*, and that Hart’s appeal to consequentialist considerations in answering the first question and nonconsequentialist considerations in answering the second and third creates a dialectic instability in his view. Others have endorsed the disaggregation strategy but have argued that a normatively prior to the question of why should we punish, which Hart called punishment’s general justifying aim, is the question of whether the practice itself is morally permissible.¹ As these scholars point out, to demonstrate that there is good reason to *X* does not yet establish that it is permissible to *X*. On this view, defenders of punishment must first explain why a community’s political authority can be justified in imposing on them the sort of harms that are characteristic of punishment,

¹ C.f., K. G. Armstrong, “The Retributivist Hits Back,” in H. B. Acton, ed., *The Philosophy of Punishment* (London: Macmillan, 1969), p. 141; David Dolinko, “Some Thoughts about Retributivism,” *Ethics* 101:3 (1991): 537-59, on pp. 539-41; and Matt Matravers, *Justice and Punishment* (Oxford, U.K.: Oxford University Press, 2000). Matravers writes that “punishment theory must concern itself with the morality of attaching the threat of sanctions to rules (as well [as] the morality of imposing those sanctions on particular people). And whilst it seems plausible to think that the point of threatening sanctions must have something to do with preventing offending . . . , that is not the same as arguing that preventing offending through the threat and imposing of sanctions is morally permissible” p. 7. Note also the distinction between the moral permissibility question and the “whom may we punish?” question: The answer to the latter question might be, e.g., “only those culpable for criminal wrongdoing,” but this answer would, in itself, say nothing about why *punishment*, i.e., the infliction of intentional harm, is a morally permissible response to criminal wrongdoing.

harms that would be clearly impermissible if inflicted on law abiders. As David Boonin writes:

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?²

I refer to this throughout as the question of punishment's in-principle moral permissibility. If the institution is not in principle permissible, then appeal to whatever positive aims will be insufficient to justify it.

In this paper, I offer a defense of punishment's in-principle permissibility. My account is a version of the fair play view, according to which, briefly, the permissibility of punishment derives from reciprocal obligations shared by members of a political community, here understood as a mutually beneficial, cooperative social venture. Mine is a nonstandard fair play account, however, in that most fair play accounts aspire to offer unified theories of punishment — that is, they employ considerations of fair play to ground not only punishment's in-principle permissibility, but also its positive aim as well as sentencing guidance. By contrast, my fair play view is more modest; it seeks only to provide an answer to the permissibility question. I contend that in this context, modesty is a virtue. Because my account offers only an answer to the in-principle permissibility question, but not to the positive aim question or to questions of how we may punish, it avoids certain powerful objections that have been raised against standard articulations of

² Boonin, *The Problem of Punishment* (Cambridge, U.K.: Cambridge University Press, 2008), p. 1.

the fair play view. What's more, as I argue below, my focus on only the permissibility question is not *ad hoc*; to the contrary, a closer examination of the fair play view's evolution from a theory of political obligation to a defense of punishment indicates that there are good reasons to expect that it is suitable as an answer only to the permissibility question. Punishment's positive aim and the constraints on how it is administered in particular cases must be based on distinct considerations.

It's worth emphasizing at the outset that the challenge with which I am concerned here is not that this or that sort of punishment (or punishment for violation of these or those laws, or within this or that political system) is impermissible, but rather that the practice of punishing *per se* is impermissible. If this objection is correct, then all punishment will be ruled out from the start. By the same token, even if, as I argue below, considerations of fair play can ground a satisfactory answer to the in-principle permissibility question, actual inflictions of such harm could nevertheless be impermissible — e.g., as a response to unjust or unreasonable laws, or when inflicted to an excessive degree or in an inhumane manner, etc. My concern in this chapter is thus not to provide a complete justification for punishment, but rather to establish that, and explain why, punishment is in principle a permissible response to criminal violations.

In section I, I examine the standard articulation of the fair play view. I consider how the view, first offered as an account of political obligation, has been extended to justify punishment, and why this justification ultimately fails. In section II, I develop my alternative version of the fair play view, on which the defense of punishment's in-principle permissibility follows more straightforwardly from fair play's answer to the

political obligation question. I contend that my version of the view fares better than standard articulations on a number of counts. Finally, in section III, I consider certain objections to my view. As will become clear, these objections essentially are objections to the fair play view of political obligation — and although I do not attempt here to offer a full defense of this broader view, I do at least aim to address particular concerns that may emerge from my inclusion of punishment among the rules to which we have a moral obligation to comply.

I. The standard fair play account and its drawbacks

According to the fair play account, a political community can be understood as a cooperative venture in which each member benefits when there is general compliance with the rules governing the venture. The fact that each member benefits from the compliance of other members generates an obligation to reciprocate by similarly complying. Thus Hart wrote,

[W]hen a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission.³

As espoused by Hart, the fair play view grounded a reciprocal obligation to comply with the rules of a mutually beneficial political community. It said nothing about

³ H. L. A. Hart, “Are There Any Natural Rights?” *The Philosophical Review* 64:4 (April 1955): 175-91, reprinted in Jeremy Waldron, ed., *Theories of Rights* (Oxford, U.K.: Oxford University Press, 1984), on p. 85. For a similar articulation of this view, see John Rawls, “Legal Obligation and the Duty of Fair Play.” In *Law and Philosophy: A Symposium*, ed. S. Hook (New York City: New York University Press, 1964), pp. 9-10.

what would be a justified response to those who failed to meet this obligation. Several theorists of punishment, however, have used the principle of fair play as their foundation in developing a defense of the practice of punishment. The crucial claim for extending the fair play view to justify punishment is that when a member of the community chooses not to comply with the community's laws, she takes an *unfair advantage* relative to her fellow community members. That is, she unfairly benefits twice: Like everyone, she reaps the benefits that general compliance with the law makes possible, but she additionally benefits in that she, unlike her fellow community members, doesn't constrain her behavior in compliance. Typically, then, on fair play accounts the offender is portrayed as a free rider, and punishment is defended as a means of removing the offender's unfair advantage by imposing a burden on the offender proportionate to the additional benefit she unfairly gained through her crime. As Herbert Morris writes, "Justice — that is, punishing such individuals — restores the equilibrium of benefits and burdens by taking from the individual what he owes, that is, exacting the debt."⁴

The standard articulation of the fair play view of punishment is inadequate in two key respects: The first is that the fair play view often misconstrues what makes a criminal act worthy of punishment, or as R. A. Duff writes, "it offers a distorted picture of the

⁴ Herbert Morris, "Persons and Punishment," *Monist* 52 (1968): 475-501; on p. 478. For other notable elaborations of the fair play view, see Jeffrie Murphy, "Three Mistakes about Retributivism," *Analysis* 31 (1971): 166-9; Michael Davis, "How to Make the Punishment Fit the Crime," *Ethics* 93 (1983): 726-52, and "Criminal Desert and Unfair Advantage: What's the Connection?" *Law and Philosophy* 12 (1993): 133-56; George Sher, *Desert* (Princeton, N.J.: Princeton University Press, 1989); and Richard Dagger, "Playing Fair with Punishment," *Ethics* 103 (April 1993): 473-88.

punishment-deserving character of crime.”⁵ That is, we tend to think that a person who has, for instance, tortured someone should be punished not because she has gained an unfair advantage over other members of the community generally, but rather because of the heinous moral wrong she has committed against her victim. In other words, we do not typically think of serious *mala in se* crimes such as torture, murder, or rape as primarily matters of free riding.

The second deficiency of standard fair play accounts involves the specification of the offender’s unfair benefit. Put simply, there doesn’t seem to be any advantage that an offender gains, in proportion with the seriousness of her crime, relative to community members generally.⁶ Here I briefly consider three suggestions for this unfair advantage.

⁵ R. A. Duff, *Punishment, Communication, and Community* (Oxford, U.K.: Oxford University Press, 2001), p. 22.

⁶ It’s worth distinguishing two types of benefits that are *not* intended by any of the fair play advocates. First, the benefit gained by an offender is not characterized as a moral benefit. M. Margaret Falls criticizes the fair play view as being incompatible with “the tradition that says willing the moral good is the highest human good and therefore doing evil harms the evildoer.” M. Margaret Falls, “Retribution, Reciprocity, and Respect for Persons,” *Law and Philosophy* 6:1 (1987): 25-51, on p. 31. I think it’s fair to say, however, that the fair play view operates within a tradition that recognizes a distinction between moral and prudential benefit, which believes that the latter does not necessarily collapse into the former, and which holds that an offender gains some *prudential* benefit through her crime. (Of course, *if* it’s true that what is prudentially good for us reduces to what is morally good for us, or even if any ostensible prudential benefit a criminal gains would be outweighed by the moral harm so that the criminal should be understood as harming herself all things considered, this will only support my conclusion below: that the fair play view cannot demonstrate an advantage that the criminal unfairly gains over others that is appropriately removed by punishment.) Second, the benefit gained by the criminal is explicitly not characterized as the material spoils of her crime. Thus, the relevant benefit unfairly gained by, say, the burglar is not the actual money or property that she steals, nor is the tax evader’s relevant benefit the tax money she doesn’t pay. If the benefit were characterized as the material gain from the crime, then removing this benefit would seem to be a matter merely of requiring the offender to *compensate* her victim(s); punishment, understood as the intentional imposition of hard treatment, would not seem necessary. For fair play defenders of punishment, therefore, it is crucial that the unfairly gained benefit is something distinct from the ill-gotten material gains.

One option is that the offender gains freedom from the burden of self-constraint that others accept in complying with the particular law that the offender violates. If so, then the appropriate severity of punishment will be proportionate to the burden others feel in complying with that law. But compliance with laws is often no real burden for most citizens. In fact, compliance with prohibitions on egregious offenses (murder, assault, etc.) typically is less burdensome than is compliance with prohibitions on lesser crimes (tax evasion, jaywalking, etc.), given that we may be more often tempted to commit the lesser crimes. Most of us are typically not tempted to commit murder or assault anyway, whereas we may feel comparatively more tempted, on occasion, to cheat on our taxes, jaywalk, etc. Thus relatively less serious violations will often appear to merit relatively more severe punishments, a deeply counterintuitive conclusion.

Instead, perhaps the offender gains freedom from the burden of compliance with the law in general. This general compliance, Richard Dagger writes, is a genuine burden: “there are times for almost all of us when we would like to have the best of both worlds — that is, the freedom we enjoy under the rule of law plus freedom from the burden of obeying laws.”⁷ This route, however, appears to lead to the objection that all offenses become, for the purposes of punishment, the same offense. Both the murderer and the tax cheat have failed to comply with the rule of law generally while benefiting from the general compliance of others. If the punishable offense is the same, however, then the two cases appear to warrant equal punishments, and again, this strikes most of us as

⁷ Dagger, “Playing Fair with Punishment,” p. 483.

counterintuitive.⁸ Thus the particular- and general-compliance views appear unsatisfactory.

A third account, by George Sher, holds that the offender gains an extra measure of freedom, not from the burden of self-constraint, but rather “from the demands of the prohibition he violates.”⁹ Sher writes: “Because others take that prohibition seriously, they lack a similar liberty. And as the strength of the prohibition increases, so too does the freedom from it which its violation entails.”¹⁰ Although Sher’s account appears to avoid the counterintuitive sentencing implications that beset the previous two views, his account faces its own problems. Specifically, it’s not clear in what sense the offender, by committing the crime, *gains* freedom from the moral prohibition. As David Dolinko points out, the criminal does not so much *gain* freedom as *exhibit* a freedom he already had — “he must have been ‘free’ from the prohibition even before his lawless act (or he could not have committed it!), and presumably, many law-abiding citizens are equally ‘free’ (in this sense) to violate the prohibition.”¹¹ Ultimately, Sher’s account fares no better than the particular- and general-compliance views.

If the relevant benefit that an offender unfairly gains is not any of these types of freedom, then what else might constitute her unfair advantage? There certainly may be other things that an offender gains through her commission of a crime. For instance, Jean

⁸ This objection is pressed by Boonin, *The Problem of Punishment*, pp. 125-6.

⁹ Sher, *Desert*, p. 82.

¹⁰ *Ibid.*

¹¹ David Dolinko, “Some Thoughts about Retributivism,” *Ethics* 101:3 (April 1991): 537-59, on p. 547.

Hampton's expressive retributivism holds that wrongful acts "convey — and work to effect — the wrongdoer's superior importance relative to the victim understood as an individual or as a class of individuals."¹² Thus we might follow Hampton in regarding the offender as gaining a sense of, and perhaps a realization of, relative superiority.

Hampton's account is explicitly not a fair play account, however, as this ostensible benefit to the offender, the superiority that the wrongful act expresses and seeks to manifest, is a superiority to the particular victim(s), not to other, law-abiding community members generally. For purposes of the fair play account, the punishable benefit must be something that the offender gains relative to the community in general, as a result of others' compliance and her own noncompliance; that is, the benefit must be a result of her free riding. And insofar as punishment is justified as a means of removing the unfair advantage, this advantage must either be commensurate with the gravity of the crime or else risk running afoul of our deeply held intuitions regarding proportionality of punishment. Unfortunately for advocates of the standard fair play account, there just doesn't appear to be any unfair advantage that all criminals gain, in proportion to the seriousness of their crimes, over other, law-abiding community members generally.

Dagger has rearticulated, and further developed, his view in a recent article titled "Punishment and Fair Play."¹³ Here he maintains that all crimes are indeed crimes of unfairness, but he contends that they may be unfair not only in the sense of yielding

¹² Jean Hampton, *The Intrinsic Worth of Persons: Contractarianism in Moral and Political Philosophy*, ed. Daniel Farnham (New York City: Cambridge University Press, 2007), p. 131. See also Hampton, "A New Theory of Retribution," in *Liability and Responsibility*, eds. C. Morris and R. Frey (Cambridge: Cambridge University Press, 1991), pp. 377-414.

¹³ Richard Dagger, "Punishment as Fair Play," *Res Publica* 14 (2008): 259-75.

unfair benefits, but also in undermining the political order. If we conceive of a political community as a fair cooperative practice whose members have equal standing, Dagger contends, then “considerations of unfairness can also justify the conclusion that some offenses are more serious violations of equal standing and fair play than others.”¹⁴ For instance, he writes:

The tax evader takes advantage of many people — millions of them in many cases — but her offense typically does not make it difficult for them to continue doing their part in the cooperative practice. With the rapist, the murderer, and the batterer, however, the offender has done something that makes it difficult or even impossible for his victim to contribute further to the ongoing cooperation. He has offended against the interests and integrity of his victim, to be sure, but he has also offended against the requirements of a society based on fair play, and his offense is thus a more serious crime of unfairness than the tax evader’s.¹⁵

There are, I believe, two significant problems with this argument. First, it’s not clear that the rapist does make it more difficult for his victim than the tax evader makes it for her victims to contribute further to the ongoing cooperation. As Dagger has (rightly) characterized it, the relevant sense of cooperation here is cooperation in complying with the law in general. When others exercise general compliance with the law, and when I benefit from their compliance, then I have an obligation similarly to participate in the cooperative venture (i.e., to reciprocate) by complying with the law. But although rape clearly is a more egregious violation than tax evasion, it’s not clear that one way in which

¹⁴ Ibid., p. 270.

¹⁵ Ibid.

it is more egregious is that the rape victim's ability to "further the ongoing cooperation" by accepting the burden of compliance with the law is especially diminished.

Second, even if the more serious crime does more severely threaten its victim's ability to contribute to the fair cooperative venture, this does not demonstrate that such a crime is a more serious crime *of unfairness* than the less serious crime. Put more simply, an act may undermine fairness without itself being unfair.¹⁶ On the standard fair play view's characterization, crimes are unfair in the sense of free riding. Offenders accept the benefits made possible by the general compliance of others with the law, and then they choose not to reciprocate. It's just not clear, however, that by more seriously undermining the fair political order, an offender has therefore been more of a free rider. In fact, given that the offender (like everyone) benefits from the cooperative venture, then to the extent that she offends against the cooperative venture (by undermining the victim's ability to contribute to it), she is actually more likely to harm rather than benefit herself. Free riders can only ride freely when the practice from which they draw benefits, but to which they do not contribute, thrives. Thus it is unclear how, by more seriously offending against the

¹⁶ The distinction I have in mind here, between undermining fairness and being unfair, is essentially one made by Philip Pettit in his article "Consequentialism," in *Consequentialism*, ed. Stephen Darwall (Malden, Mass.: Blackwell Publishing, 2003), pp. 95-107, esp. p. 97. In discussing the difference between consequentialists and nonconsequentialists, Pettit points out two distinct ways in which we may respond to whatever we value: We may promote it, or we may honor it. What's more, promoting what we value doesn't necessarily imply honoring it, and vice versa. The converse is the distinction I have in mind: We may undermine some value, or we may violate it.

cooperative venture, the offender would more egregiously free-ride than in cases in which she less seriously offends against the cooperative venture.¹⁷

Ultimately, the fair play view is unable to provide a plausible, univocal account of punishment that grounds not only its in-principle permissibility but also its positive aim and sentencing guidance in particular cases. Traditional fair play articulations provide the wrong answer, at least in many cases, to the question “why should we punish this crime?” In addition, they are unable plausibly to specify any benefit that an offender unfairly gains, in proportion to the seriousness of her crime, over law-abiding community members generally. And although I believe Dagger’s recent fair play defense is a significant improvement over traditional accounts, it is not ultimately an integrated account. Rather, it implicitly appeals both to the traditional, deontological conception of fair play as well as the consequentialist aim of preserving a fair political order. As will become clear in the next section, however, I do not regard fair play’s inability to answer all the questions of punishment (i.e., its inability to ground a unified account) as a liability. Rather, on the account I propose, there is good reason to expect that the fair play view will ground only the in-principle permissibility of punishment, and that punishment’s aim and guidance regarding appropriate sentencing will require appeal to distinct considerations.

¹⁷ Consider, by analogy, which is the greater violation of fair play (i.e., the greater instance of free riding): the citizen who avoids paying taxes but nevertheless reaps benefits from the flourishing tax system, or the citizen who actively works to destroy the institution of taxation itself.

II. An alternative fair play account

Although I have argued against prominent fair play accounts of punishment, I nevertheless find something intuitively appealing about the fair play account of political obligation espoused by Hart and Rawls. In beginning to set out my own fair play account of punishment, then, I believe Hart's concise statement of the fair play principle bears repeating:

[W]hen a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission.¹⁸

So, as a member of a cooperative enterprise, if I benefit from others' playing by the rules, then I should play by the rules as well. But which rules? The rules most frequently appealed to by fair play accounts of political obligation and of punishment are the political community's criminal statutes, the laws prohibiting, say, murder, theft, tax evasion, drug trafficking, etc. Two features of such rules are relevant for present purposes: First, they are the sort of rules with which we can *comply* (by not murdering or stealing, by paying our taxes, etc.). Second, general compliance with these rules provides a significant *benefit*. These are the two salient features for generating the fair play obligation: If general compliance with a rule is beneficial to me, then I have an obligation of fairness similarly to comply.¹⁹ In this section, I defend the claim that the rule

¹⁸ *Supra*. note 9.

¹⁹ Some have objected that it is *acceptance* of benefits, not merely receipt of benefits, that can generate a fair play obligation. I consider this objection in section III.

instituting punishment as a response to crimes is itself one of the rules with which we have an obligation of fairness to comply. Thus on my account, punishment is in-principle permissible not because it removes some benefit offenders have unfairly gained relative to law abiders. Rather, punishment is permissible because the rule instituting punishment as a response to crimes is itself one of those rules with which we, who benefit from general compliance with the rule, have a fair play obligation to comply. To defend this claim, I need to say more about how I understand the rule instituting punishment as a response to crime. In particular, I need to establish that this is the sort of rule with which we can comply, and also that general compliance with this rule yields significant benefits.

The rule instituting punishment is a rule of remediation. Whereas criminal laws tell us things such as “don’t commit murder” or “pay your taxes,” the rule instituting punishment tells what is to be done when community members violate these laws. There is nothing in the conception of a criminal law that entails that the law must be backed by punishment.²⁰ Indeed, other forms of response to the violation of such laws have been suggested: nonpunitive censure or restitution, for example. Thus if punishment is to be the response to criminal violations, this will be because it has been so designated by some rule of the political community. Essentially, this rule takes the form of a conditional, and

²⁰ Hart appears to have disagreed on this point, as he indicates in various passages that, as a conceptual matter, criminal laws must be backed by physical sanctions, i.e., punishment (see, e.g., Hart, *The Concept of Law* 2d ed. (Oxford: Oxford University Press, 1994), pp. 34-5, 86). If this is right, so that criminal laws without punishment are not really criminal *laws*, then those who would endorse the abolition of punishment will face the unenviable task of also defending the abolition of criminal laws altogether. As I have indicated, however, I reject the view that criminal law entails punishment. Thus on my view, even if criminal statutes are themselves justified, the proposition that punishment is an appropriate mode of response to violations of these statutes nevertheless requires its own defense. I am grateful to Larry May for raising this point to me.

it says that if you commit a crime, then you will be subject to having your liberties restricted in ways to which law abiders are not subject. The question, for present purposes, is whether this is a rule with which we can comply, and if so, whether general compliance with it yields significant benefits.

On first blush, it may seem that the suggested rule is not one with which we can comply. That is, it might appear to be what Hart, in his *The Concept of Law*, called a secondary rule rather than a primary rule. Put simply, Hart characterized primary rules as imposing duties or obligations, and secondary rules as conferring powers.²¹ Unlike primary rules such as “don’t commit murder,” the rule instituting punishment may appear less the issuance of a command than an instruction to legal authorities as to what may be done to us if we violate rules of the first type. I contend, however, that the rule instituting punishment is not solely an instruction to legal authorities. A significant element of the institution of punishment is that it communicates to, and indeed imposes obligations on, citizens themselves.

To construe the rule instituting punishment merely as an instruction to officials, e.g., “punish those who violate criminal statutes,” is to overlook an important communicative element of punishment. The institution of punishment communicates to citizens generally that the community condemns certain actions as morally wrong — condemns them so strongly that it is willing to impose hard treatment on those who

²¹ Hart, *The Concept of Law*, esp. pp. 79-99.

commit such actions.²² In my view, the central benefit of this threat of hard treatment is its role in reducing the frequency of violations of community members' security and well-being. The institution of punishment, then, communicates the central importance that the community places on protecting its members. Similarly, it asks its members, *as* community members, to share in this commitment. And insofar as punishment itself plays a key role in securing these important aims, the community asks its members to comply with this institution.

Still, we might wonder whether — and if so, how — one could comply with the rule instituting punishment? We tend to think of punishment, after all, as coercively imposed on offenders who may be no more than passive recipients. I contend, however, that the rule instituting punishment as a response to crimes is one with which we can comply (or fail to comply). As I have indicated, I believe the rule instituting punishment does communicate to community members, and it is a rule of remediation: It tells us that if we commit some crime, then we should accept being subject to punishment as a response. One way to comply, then, would be to constrain one's behavior so as to avoid being liable to punishment; another way to comply would be, if one has committed a criminal offense, to accept the prescribed punishment. Conversely, one could fail to

²² For discussions of the communicative aspect of punishment, see, e.g., Joel Feinberg, "The Expressive Function of Punishment," in *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton (N.J.: Princeton University Press, 1970); Jean Hampton, *The Intrinsic Worth of Persons: Contractarianism in Moral and Political Philosophy*, ed. Daniel Farnham (New York City: Cambridge University Press, 2007); and Duff, *Punishment, Communication, and Community*. Note that, unlike Hampton and Duff, my view is not that this communicative aspect itself grounds the permissibility of punishment. But I do accept that an aspect of punishment is communicative, and as I discuss, part of this communication is to ask something of all community members, law abiders and offenders alike — viz., that if they don't comply with the community's criminal statutes, then they should accept being subject to punishment.

comply with this rule by committing a crime and attempting to evade apprehension (and subsequent punishment). In other words, if the rule is stated as the conditional “if you commit a crime, then you are subject to punishment,” then we can comply either by not committing crimes, thus rendering the antecedent false (and the conditional itself trivially true), or by accepting being subject to punishment, so that the consequent (and thus the conditional) is rendered true. Both forms of compliance warrant further explanation.

First, we can comply by not committing crimes. A likely objection here is that when we refrain from committing crimes, we comply not with the proposed rule of punishment, but rather with the rule prohibiting the particular crime. I offer two responses: In one sense, when we refrain from committing crimes, we comply with both rules. We comply straightforwardly with the rule that says “do not murder,” and we comply with the remedial rule (“if you commit a crime, then you are subject to punishment”) by falsifying the antecedent and thus rendering the conditional trivially true.

In another sense, whether we can be understood as complying with the rule instituting punishment may depend on our reasons for compliance. Antony Duff argues that the criminal law of a liberal polity is best understood as offering moral reasons, not prudential reasons, for compliance.²³ I agree that the criminal law should offer moral reasons. But unlike Duff, I believe the institution of punishment often permissibly offers

²³ Duff, *Punishment, Communication, and Community*, pp. 78-79. He writes, “[T]he reasons that citizens have to refrain from [criminal] conduct, the reasons to which the law refers and on which it depends, are precisely the moral reasons that make such conduct wrong.”

prudential reasons for compliance.²⁴ When we comply with the law because we accept our community's moral condemnation of the criminal act, we do not need the threat of punishment to motivate our compliance. But inevitably, there will be times when the moral appeal of the criminal statute itself is not sufficient to motivate us. In these cases, the threat posed by the institution of punishment may provide a prudential incentive to do what the moral reason was not sufficient to persuade us to do. I suggest that in cases when we constrain our behavior not because of the moral reasons offered by the law itself but because of the prudential reason presented by the threat of punishment, we may be said to comply with the remedial rule: if you commit a crime, then you will be subject to punishment.

A second way to comply with the proposed rule is by accepting, if we do commit some crime(s), being subject to punishment. In my view, this rules out attempting to evade apprehension or falsely representing oneself as innocent once one has been apprehended (though it does not prohibit appealing to what one may believe are genuinely mitigating circumstances). Also, the punishment itself may require an active response from the offender. She may be required to pay a fine, to appear for community service, or to meet with her probation officer. Even incarceration may require the offender's active participation: For instance, courts will in some cases require the convicted person to report to prison on a certain date to serve her term.²⁵ Thus those who

²⁴ On this point, I am in general agreement with Andrew von Hirsch's view. See, e.g., von Hirsch, *Censure and Sanctions* (Oxford, U.K.: Clarendon Press, 1993).

²⁵ I thank Antony Duff for suggesting this point to me. More generally, I follow Duff in thinking that respect for offenders as autonomous moral agents, and as still members of the political community, requires

commit crimes, although they fail to comply with the particular rules prohibiting the crimes they commit, may still comply with the rule instituting punishment if they accept being subject to punishment, in ways such as I've suggested here.

I conclude, then, that the rule instituting punishment is one with which we can comply. The second question is whether general compliance with this rule yields significant benefits. I doubt that we receive significant benefit from general compliance of the second sort (accepting punishment if one has committed a crime), simply because most people who have committed crimes typically do not accept punishment in ways such as I suggested.²⁶ Many people do comply, however, in the first way: by not committing crimes. We should ask, then, whether general compliance of this first sort yields significant benefits to community members. If so, then insofar as I reap these benefits of general compliance, I have an obligation to comply as well.

To deny that general compliance with the rule instituting punishment yields substantial benefits, one would need to demonstrate that the protections afforded to citizens by the rule of law would not be significantly undermined if punishment were abolished (perhaps to be replaced by some alternative, such as public censure or restitution). Such an argument would thus need to refute the intuitively compelling and widely accepted claim that punishment has a substantial deterrent effect.²⁷ If general

that we engage them as active participants in their punishment, rather than passive recipients (see, e.g., Duff, *Punishment, Communication, and Community*).

²⁶ Although perhaps some, such as already repentant offenders, may do so.

²⁷ Engaging with the important empirical debate about whether in fact punishment yields significant deterrent effects is beyond the scope of this chapter. I note, however, that if there are no considerable deterrent impacts from maintaining the institution of punishment, then the fair play account I have

compliance with the rule instituting punishment does provide significant benefits, then it falls within the scope of those rules to which the fair play account grounds an obligation of compliance.

If the rule that says criminal violations are subject to punishment is among those rules with which we are obliged to comply, then the remainder of my argument for punishment's in-principle permissibility follows fairly straightforwardly. Because having a moral obligation to *X* implies not having a moral right not to *X*, it follows that none of us has a moral right not to comply with this rule. Because, as I indicated earlier, compliance for a lawbreaker requires accepting one's punishment, it follows that no one has a right not to be punished if she violates some criminal statute. The practice of punishment in general, then, is not a violation of criminals' rights.

To briefly sum up my argument to this point: Punishment is in principle morally permissible not because it removes some benefit(s) that an offender unfairly gains in failing to meet her obligation to play by the community's legal rules; rather, it is permissible because punishment as a response to crime is itself one of the rules with which the offender, like all those who benefit from mutual compliance with the rules, is obliged to comply. My fair play account of punishment *just is* the fair play account of political obligation, along with the recognition that the rules of punishment are among

developed here would be unable to ground its in-principle moral permissibility. This fits with my own intuition, however, as I believe that punishment's yielding some significant deterrent benefit is a necessary (albeit not sufficient) condition of the institution's moral permissibility. In what follows, however, I assume (in keeping with the predominant view) that the institution of punishment — or more specifically, general compliance with the rule instituting punishment — does yield significant benefits to members of the political community.

those rules of the community that, when community members generally comply with them, are mutually beneficial. In section III, I address what I take to be the most powerful objections against the fair play view of political obligation. First, however, in the remainder of this section, I want to consider a number of advantages to my view in comparison with the standard fair play account of punishment.

Most importantly, my view fares better against the two objections to the traditional fair play account of punishment that I discussed earlier. Consider the first objection, that the traditional fair play account misconstrues why certain crimes merit punishment. Murder, that is, does not seem to be centrally, if at all, a crime of free-riding on other community members generally. My version of the fair play account, by contrast, grounds only punishment's in-principle permissibility. It does not follow from my account that fair play considerations also supply the positive aim of punishment; thus my account does not imply, for instance, that the murderer should be punished because she was a free rider on members of the community generally. In my view, the central benefit of the rule establishing punishment is that it gives citizens compelling reasons to comply with the laws — i.e., it acts as a general deterrent — and thus helps to ensure the safety and security of community members. Even if deterrence represents the reason we should want an institution of punishment, however, the institution's in-principle permissibility will stem, on the fair play view, from the fact that an offender (like everyone else) reaps the benefits of deterrence as a result of the general compliance with the rule establishing punishment — thus the offender has an obligation of fairness similarly to comply.

The second objection noted to the traditional articulation of the fair play view is that there seems to be no benefit that an offender unfairly gains, relative to other community members generally, through her commission of a crime. Because the traditional articulation justifies punishment as a means of removing the unfair advantage, the inability to specify such an advantage is obviously problematic. And accounts that have specified some advantage appear to generate counterintuitive sentencing guidance. Again, my account avoids this general line of criticism, as my view does not characterize punishment as removing some unfairly gained advantage. Rather on this account, punishment is permissible because the practice is among the rules of the cooperative system to which general compliance yields certain benefits — benefits that offenders, like everyone else, enjoy. This fair play account only establishes that, and explains why, punishment is in-principle permissible. It does not claim also to answer the question of what mode and degree of punishment are permissible in particular cases. This latter question will depend on distinct moral considerations.

An implication of my account, then, as I have indicated, is that it is appropriate to disaggregate various questions of punishment — in particular, (why) is punishment in principle morally permissible? what is punishment's positive aim? and what mode and degree of punishment are permissible in particular cases? — and to answer these questions by appeal to distinct considerations. Some may criticize this sort of disaggregation strategy as *ad hoc*, but on the fair play view I have suggested, this objection is unpersuasive. Again, my fair play account just is the fair play account of political obligation, with the rules of punishment recognized to be among those rules to

which the obligation of fairness extends. To require this account to ground not only punishment's in-principle permissibility but also its positive aim and sentencing guidance would be as implausible as requiring it to ground specific conclusions about the other rules of the cooperative enterprise. We don't expect that the fair play view should tell us, for instance, which acts should be required or forbidden by criminal statutes. Why then, should we expect the same fair play view to generate rules about punishment? Rather than its being *ad hoc* to distinguish the permissibility question from the other questions, on the fair play view we have good reason to expect that the answer to the permissibility question will not yield guidance regarding the other two questions.

As I have discussed, my account avoids what I take to be the two most powerful objections to the standard fair play articulation. More generally, however, my account has an advantage over any fair play account that defends punishment as a sort of appropriate remediation for violations of the fair play obligation to play by the rules of the cooperative venture. Such accounts require two substantial defenses: a defense of the fair play account of political obligation itself, and a defense of punishment as a permissible remediation for failures to meet the obligation defended in the first part. By contrast, once we recognize that the rules of punishment are among those rules to which the obligation of compliance extends, then on my account only one substantial defense is required: a defense of the fair play view of political obligation. My view is in this regard sturdier than standard fair play accounts of punishment, insofar as objections that purport to undermine my view will undermine the standard articulations as well, whereas not all objections faced by the standard articulations also threaten my account.

Essentially, on my fair play account, whether the institution of punishment is in-principle permissible will be determined by whether we have a moral obligation to comply with the rules of our political community. Thus, this defense of punishment will stand or fall according to whether the fair play account of political obligation is persuasive. In the following section, I consider certain objections to my view. I do not aim to provide a full defense of the fair play view of political obligation. I do, however, consider what I take to be among the strongest objections to this broader view, and in particular, their implications for my strategy of including *punishment* among those rules with which we have a fair play obligation to comply.

III. Objections

The first objection I want to consider involves my strategy of deriving the moral permissibility of punishment from an account of political obligation. On my account, the rules instituting legal punishment are among those with which we are reciprocally obliged to comply as members of a political community, here characterized as a mutually beneficial, cooperative enterprise. According to the objection, therefore, this strategy implies that punishment would not be morally permissible in the absence of such a political community, viz., in the state of nature. Insofar as we have intuitions that punishment would be morally permissible in the state of nature, then my account appears deficient.²⁸

²⁸ I thank Christopher Heath Wellman for raising this objection to me.

I actually have mixed intuitions about whether punishment would be permissible in the state of nature. On one hand, I'm somewhat inclined to maintain that punishment would not, perhaps could not, be permissible in such conditions. Those in a state of nature might retaliate against wrongs perpetrated against them, but it's not clear that harming in this context, even if proportionate to the wrongdoing, would constitute just punishment. Kant, for one, believed that just punishment is impossible in the state of nature, because there is no public authority to settle disputes.²⁹ If we accept the notion that punishment, to be permissible, must be imposed by a proper authority with standing to settle disputes between opposing parties, then it appears that such punishment is by definition impossible in the state of nature.

On the other hand, it seems that I might permissibly impose intentional harm on someone who has wronged me, even if there is no recognized authority to confirm that punishment in such a case is permissible. If this is true, however, I contend that such punishment would be morally permissible for roughly the same reasons that it is permissible in a political community. In the state of nature, if George steals from Kramer, and Kramer responds not only by retrieving his stolen goods but also by inflicting some sort of harm on George, then presumably this will tend to deter others who might have otherwise considered stealing from Kramer. Conversely, if Kramer didn't respond, others might take this as evidence that they could get away with similar behavior as George.

²⁹ C.f., Immanuel Kant, *The Doctrine of Right* 6:312, in the Cambridge Edition of the Works of Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary J. Gregor (Cambridge, U.K.: Cambridge University Press, 1996), p. 456. See also Helga Varden's helpful discussion of Kant's account in Varden, "Kant's Non-Voluntarist Conception of Political Obligations: Why Justice is Impossible in the State of Nature," *Kantian Review* 13:2 (2008): 1-45.

Furthermore, when Kramer punishes George, seeing this may lead Elaine to think twice not only about stealing from Kramer in the future, but about stealing from anyone. This is because Kramer's punitive response to George's stealing may cause Elaine, especially if she has witnessed others responding in similar retaliatory ways in similar circumstances, to believe that this sort of response is the sort that tends to follow attempts at stealing. All of this is just to say that general deterrence would be a significant benefit (arguably the central benefit) of punishing wrongdoing in the state of nature. As noted above, the deterrent effect will be particularly strong if those in the state of nature see wrongdoing meeting with punitive responses with some regularity. That is, if individuals begin to regard it as a sort of informal rule that wrongdoing is met with a punitive response, then they may be persuaded to comply with this rule by constraining their behavior to avoid the punishment. But if such compliance with this informal rule is beneficial to those in the state of nature, then they have an obligation of fair play to comply with it as well, either by appropriately constraining their behavior or by accepting the punitive response when they do engage in wrongdoing.

On the view I have developed here, considerations of fair play could ground the in-principle permissibility of punishment even in the absence of a formal cooperative scheme, such as a political community. Whether it actually did ground punishment's permissibility would be a matter of whether (a) individuals came to regard it as a sort of informal rule that wrongdoing is met with punishment, (b) recognition of the rule led to general compliance with it (with compliance here taking the form of choosing not to engage in wrongdoing so as to avoid punishment), and (c) general compliance with the

rule yielded significant benefits for individuals. Although I have argued that each of these requirements might hold in the state of nature, notice that the existence of a political community governed by the rule of law makes each of them much more likely. In such a political community, the rules of punishment are not merely regularities of behavior that may come to be seen as informal rules; rather, they are set out formally, so that everyone can clearly recognize them as rules of the community. The more clearly individuals recognize the rules, the more likely they will be to comply. The greater the general level of compliance, the greater will be the benefit — i.e., general deterrence — to community members. Finally, the greater the benefit community members enjoy from the compliance of others, the greater their (fair play) obligation is similarly to comply.

So to sum up my response to this first objection, punishment may be in-principle permissible in the state of nature, but if so it will be because of the same considerations of fairness that ground its permissibility in a political community. The fair play obligation to comply with the rules of punishment will be significantly stronger, however, in a political community than in the state of nature, because the benefits yielded by punishment will be comparatively greater in a political community than in the state of nature.

The state-of-nature objection charged that the fair play view is insufficient in that it can only establish punishment's permissibility in the context of a cooperative social order. A second line of objection contends that, even in the context of such a social order, the fair play view is insufficient in that it can only establish the permissibility of punishing those who *benefit* from the institution of punishment. The worry here applies to the fair play view of political obligation generally. The objection is that, although the

goods yielded by the cooperative social order may be generally beneficial, there may be individuals for whom the costs of compliance with the rules of the scheme outweigh the corresponding benefits. If so, then in such cases considerations of fair play seem ill-suited to ground obligations of compliance. In the context of the rule(s) instituting punishment of crimes, one might object that many criminals do not, all-things-considered, benefit from the existence of such an institution (or, to put it another way, that the costs outweigh the benefits). That is, we might be hard pressed to demonstrate that, say, an individual serving an extended prison term is better off than she would have been had there been no institution of punishment. In the counterfactual case, she might not enjoy the benefits that punishment yields (general deterrence, etc.), but she also would not face the hardships associated with the prison term. Thus on balance, one might argue that she would be better off if there were no institution of punishment.³⁰

Whether an individual being punished would be, on balance, better off in a society with no institution of punishment is an empirical question. It is not obvious to me that she would be better off in such circumstances. On one hand, without the prospect of legal punishment to deter others from committing crimes, the individual's own safety and security (as well as that of her loved ones) might be significantly jeopardized. On the other hand, if the person herself engaged in wrongdoing against others, the void created by the absence of a legal institution of punishment might be filled by private vengeance. This vengeance might be much more severe than the legal punishment that the offender

³⁰ I am grateful to Victor Tadros for raising this objection to me.

would face.³¹ More importantly, however, as I noted earlier, the rule instituting punishment as the response to crimes is best understood as a conditional: *if you commit a crime, then you will be subject to punishment*. The rule thus offers individuals not only the benefits of general deterrence (or perhaps special deterrence, incapacitation, retribution, etc.) but also the opportunity to constrain their behavior so as to avoid punishment. So to the prisoner who claims she has not benefited from the rule instituting punishment, we might reply: “Is it more beneficial (a) that there be a rule that helps protect you and your loved ones by deterring crimes and allows you the opportunity to avoid being punished yourself, or (b) that there be no such rule, so that harms to you and your family may go undeterred, and if you wrong others you will be subject to their vengeance?” If we thus consider the benefits individuals receive from the institution of punishment and the choice each person has to avoid punishment herself, then punishment does appear, on balance, beneficial.³²

³¹ Lisa H. Perkins cites the hazards of private vengeance to ground her justification of punishment in “Suggestion for a Justification of Punishment,” *Ethics* 81:1 (Oct. 1970): 55-61. See also John Locke, *Second Treatise of Civil Government*, ch. 2 para. 13, in *Readings in Social & Political Philosophy*, ed. Robert M. Stewart (New York: Oxford University Press, 1996), pp. 15-6; John Hospers, “Punishment, Protection, and Retaliation,” in *Justice and Punishment*, ed. J. B. Cederblom and William L. Blizek (Cambridge, Mass.: Ballinger Publishing Co., 1977), p. 35; and Nicola Lacey, *State Punishment: Political Principles and Community Values* (London: Routledge, 1988), p. 184.

³² One might press the objection by claiming that surely there may be those — mob bosses or drug lords, for instance — for whom the legal institution of punishment is not, on balance, beneficial. I actually think mob bosses and drug lords do benefit from, and actually depend on, the existence of institutions of law enforcement and punishment to preserve the social order in which they illegally operate. Nevertheless, I concede the general point that if examples can be produced of criminals who cannot be said to benefit from the institution of punishment (yet for whom we nevertheless believe punishment is permissible), then this represents a serious challenge to the fair play account. I submit that proponents of the fair play account of punishment would do well to focus on objections such as this, rather than trying to establish what sort of advantage an offender unfairly gains, relative to other community members, through the commission of her crime. I thank Julia Driver for pressing me on this point.

The previous objection contended that the fair play view is insufficient in that it can only ground the permissibility of punishing those who benefit from the institution of punishment. The final objection I want to consider charges that the fair play view is insufficient in that it can only ground the permissibility of punishing those who *accept* the benefits of this institution. It has been commonly suggested that my merely *receiving* benefits from others' compliance with the rules of a cooperative enterprise is not itself enough to generate obligations on me to reciprocate.³³ Rather, as A. John Simmons has written, what is required is that I *accept* these benefits.³⁴ With respect to certain kinds of benefits, which Simmons calls readily available, determining whether we accept them is fairly straightforward — if we seek them out and obtain them, then we have accepted them. For instance, if I request and receive “*special* protection by the police, if I fear for my life, say, or if I need my house to be watched while I’m away,” this would constitute my acceptance of a readily available benefit.³⁵ By contrast, many benefits of membership in a political community are not the sort that we seek out; rather, they are *open* benefits,

³³ Robert Nozick notably made the point that mere receipt of benefits is not sufficient to confer obligations by way of this example: Imagine you live in a neighborhood in which a group of your neighbors buys a public address system and decides to start a public entertainment program for the neighborhood (which happens to comprise 365 neighbors). Each neighbor is assigned one day per year in which she is responsible for running the PA system. “After 138 days on which each person has done his part, your day arrives. Are you obligated to take your turn? You *have* benefited from it, occasionally opening your window to listen, enjoying some music or chuckling at someone’s funny story. The other people *have* put themselves out. But must you answer the call when it is your turn to do so? As it stands, surely not.” Nozick, *Anarchy, State, and Utopia* (New York City: Basic Books, 1974), p. 93.

³⁴ See A. John Simmons, “The Principle of Fair Play,” *Philosophy & Public Affairs* 8:4 (1979), reprinted in Robert M. Stewart, ed., *Readings in Social & Political Philosophy* (New York City: Oxford University Press, 1996), pp. 66-81.

³⁵ *Ibid.*, p. 77.

which we cannot avoid, except perhaps at great inconvenience.³⁶ Examples of open benefits include police protection, national security from external threats, assurance of air- and water-quality standards, etc.

On Simmons' account, acceptance of an open benefit normally involves "taking the benefit willingly and knowingly," where this requires, at least, (a) regarding the benefit "as flowing from a cooperative scheme" rather than seeing it "as 'free' for the taking," and (b) thinking that the benefit is "worth the price we must pay for [it]," so that given a choice of taking the benefit and accepting the concurrent burdens or rejecting the benefit, we would take it.³⁷ These are fairly steep requirements on what counts as acceptance of a benefit. Not surprisingly, he concludes that many, perhaps most, citizens do not meet these criteria for acceptance of benefits. Many do not notice or think much about the benefits they receive from the political order, and many of those who do think about these benefits mistakenly undervalue them relative to the corresponding burdens — thus for Simmons they cannot be said to have accepted the open benefits in the sense necessary to confer political obligation.

George Klosko has provided what I take to be a persuasive response to this objection. Essentially, Klosko contends that acceptance, of the sort Simmons has in mind, is not necessary in some cases for open benefits to confer fair play obligations. Klosko contends that in situations where you benefit from our compliance with the rules of a cooperative venture but do not yourself comply (i.e., in free-rider situations), fairness

³⁶ Ibid., p. 76.

³⁷ Ibid., pp. 77 and 80.

demands *either* (a) that you no longer benefit, (b) that we (i.e., the rest of us) be similarly freed from the burden of compliance, or (c) that you start to comply.³⁸ Open benefits, by definition, benefit everyone — they cannot be provided generally but withheld from certain members of the community. Thus with respect to open benefits, (a) is not an option. Klosko argues that (b) also is not an option for certain open benefits, specifically those that are indispensable to the welfare of all community members.³⁹ National defense, for instance, “is essential to the well-being of *X* and all its members, [therefore] it *must* be provided. The consequences of nonprovision would be catastrophic for all concerned”⁴⁰ — to the free rider herself as well as everyone else. Thus to allow that no one has the burden of compliance (and thus to sacrifice the corresponding benefits) is not a practically viable option. With (a) and (b) unavailable as options, only (c) remains — the free rider is obliged to comply. Klosko writes:

It is difficult to imagine what Pickerel could say to the members of *X*, who have provided him with national defense, in order to justify his unwillingness to cooperate. Because the benefits are indispensable, he could not say that he does not want them. Nor could he distinguish himself from the other *X*-ites because he has not sought the benefits out. Because of the nature of national defense, *none* of the *X*-ites have pursued them. The *X*-ites can be presumed to differ from Pickerel in their willing acceptance of the scheme’s burdens. But Pickerel’s unwillingness to participate is difficult to defend. Unless there is some morally relevant difference between Pickerel and the members of *X*, his refusal to cooperate must be interpreted

³⁸ C.f., Klosko, *The Principle of Fairness and Political Obligation* (Lanham, Md.: Rowman & Littlefield Publishers, Inc., 1992), p. 35.

³⁹ *Ibid.*, pp. 39-54, esp. p. 43.

⁴⁰ *Ibid.*, p. 43.

simply as a desire to profit from their labor without doing his fair share, and so as a clear instance of free riding.⁴¹

Klosko thus concludes that we may be obliged to comply with rules that provide us with open, indispensable benefits even if we have not accepted these benefits in the sense Simmons requires. Note, however, that on this account, fair play *only* grounds obligations to comply with the rules that provide open and indispensable benefits.⁴² The relevant question for present purposes, then, is whether the benefits provided by the institution of punishment are open and indispensable. I contend that they are both. As I indicated earlier, I believe the central benefit of the institution of punishment is that it gives genuine bindingness to the rule of law by providing significant incentives not to violate legal rules (i.e., through general deterrence). In this way, the institution of punishment plays a crucial role in ensuring the security of community members. If I am right, then this seems fairly clearly to be an open benefit. Receiving this benefit does not require actively seeking it, and in fact it's not clear how we might refuse this benefit.

⁴¹ Ibid., p. 42. One might object to Klosko's claim here by pointing out that a community member could sincerely (albeit unwisely) claim not to want the benefits provided by national defense (or for our purposes, punishment), or at least not to want them enough to make cooperating worthwhile. This is a fair point. As Klosko's phrase "must be interpreted" indicates, there is an intractable problem in such cases of determining whether the noncooperator is genuine about being willing to forego the benefits. Because the benefits are open and also indispensable to the community generally, there is no practical possibility of actually testing whether the noncooperator's claim is sincere (and of course, it is reasonable to assume that the noncooperator is aware that this is so). Klosko indicates, and I'm inclined to agree, that we thus have good reason in such cases to suspect that the noncooperator's claim is disingenuous. I recognize, however, that if a noncooperator could be determined to be sincere in her willingness to forego the benefits of the institution of punishment, then the fair play account would be hard-pressed to justify punishment in such a case. As I suggested in response to the previous objection, I believe this is the sort of challenge with which those sympathetic to a fair play defense of punishment should concern themselves, rather than questions surrounding the identification and removal of unfair advantages gained through criminal offenses.

⁴² Klosko has argued elsewhere that we nevertheless have political obligations with respect to other rules, and that these obligations are grounded in distinct principles. See his *Political Obligations* (Oxford, U.K.: Oxford University Press, 2005), esp. chapter 5. Analyzing Klosko's broader account is beyond the scope of this essay.

The benefit provided by punishment is also, I believe, indispensable. In describing indispensable open benefits, Klosko writes that this class of benefits is likely quite small; however, he maintains that it comprises, at least, goods necessary to protecting the physical security of community members, such as national defense, protection from a hostile environment, provisions for satisfying basic bodily needs, and notably, *law and order*. “That we all need the public goods just mentioned regardless of whatever else we need is a fundamental assumption of liberal political theory.”⁴³ In particular, the fundamental importance of security has been widely recognized by liberal political theorists. As John Stuart Mill pointed out, security is a requirement for the enjoyment of virtually all other goods:

... but security no human being can possibly do without; on it we depend for all our immunity from evil, and for the whole value of all and every good, beyond the passing moment, since nothing but the gratification of the instant could be of any worth to us, if we could be deprived of anything the next instant by whoever was momentarily stronger than ourselves.⁴⁴

As Mill recognized, whatever the things are that matter to us — whether these be possessions, projects, relationships, or whatever — these things will typically have value for us insofar as we can be secure in their pursuit or enjoyment. We buy things, and we count on their not being stolen or destroyed by others; we travel, and we count on the fact

⁴³ Klosko, *The Principle of Fairness and Political Obligation*, p. 40.

⁴⁴ John Stuart Mill, *Utilitarianism*, 2d ed., ed. George Sher (Indianapolis, Ind.: Hackett Publishing, 2001), p. 54. Similarly, Klosko emphasizes the importance of security in describing “law and order” as “the advantages of a secure, protected environment, which provides one with security of the person and the realistic expectation of similar security in the future that allows one to plan ahead. According to major liberal theorists, e.g., Locke, the absence of law and order in this sense makes life in the state of nature to some degree intolerable,” *The Principle of Fairness and Political Obligation*, n30, p. 59.

that those around us will drive responsibly; we work to earn a living, and we count on the fact that our employers will not take advantage of us. The rule of law plays a crucial role in ensuring the security of all community members, and the institution of punishment plays a crucial role in ensuring that the rule of law genuinely binds.⁴⁵ Thus I conclude that the institution of punishment provides an indispensable open benefit. As such, it grounds a fair play obligation of compliance even for those who have not met Simmons' standards for acceptance of the benefit.

This is not to say that the institution of punishment will therefore be permissible no matter what punishments it prescribes in particular cases. As I have indicated from the outset, the defense I have offered here is only of punishment's in-principle moral permissibility. That is, my argument has been that punishment *per se*, that is, the intentional infliction of harm on criminal wrongdoers, does not in itself constitute a violation of offenders' rights. Particular instances of punishment, however, may still be morally impermissible all things considered if they fail to treat offenders with the respect to which they are entitled as moral persons. Again, I take it that my account is similar in this respect to fair play accounts of political obligation generally. In other words, if considerations of fair play ground an obligation to comply with criminal statutes, this

⁴⁵ My claim here rests on two empirical claims, either of which might be challenged. One might contend either that the security of community members does not depend on the rule of law, or alternatively that the rule of law's bindingness does not depend on the institution of punishment. Thus even if security is understood to be an indispensable benefit, one might argue that punishment is not an indispensable means to achieving that benefit. Arguing for these empirical claims is beyond the scope of this paper, but I concede that if it could be shown convincingly that the security of community members could be ensured as (or more) effectively by means other than a system of laws backed by punishment, then the case I offer here for punishment's in-principle permissibility would be correspondingly undermined. I am grateful to R. A. Duff for pushing me on this point.

surely is only a presumptive obligation. Such accounts would not ground (nor purport to ground) an absolute obligation of compliance irrespective of the content of the statutes. In fact, any plausible account of political obligation, be it grounded in considerations of fair play, tacit consent, natural duties of justice, or whatever, will allow that in certain cases we may be permitted, perhaps even required, to violate unjust laws (perhaps through civil disobedience, or in extreme cases, even outright revolution). Nevertheless, there is a presumptive moral obligation to comply with a community's laws — and relevant for present purposes, there is a presumptive moral obligation to comply with the rule according to which one is subject to punishment when one has violated some criminal statute. Thus the institution of punishment is in principle morally permissible.

VI. Conclusion

In this chapter, I have aimed to provide a more plausible version of the fair play justification of punishment, one that follows more straightforwardly from the fair play account of political obligation and also avoids the objections typically leveled against fair play defenses of punishment. The merits of my account could be evaluated in a couple of ways: First, we could ask whether, from within the perspective of the fair play view, my account provides a more plausible route to grounding the permissibility of punishment than do the standard articulations of the view. Second, we could ask whether the fair play view itself is plausible.

Although I am obviously sympathetic to the fair play view itself (that is, to the fair play view of political obligation), I have in this paper offered only a brief defense of

this view against what I take to be the most powerful objections raised against it. A full defense of the view is well beyond the scope of this paper. For others who would endorse the fair play account of punishment, however, I suggest that the political obligation question should take center stage. If the fair play view of political obligation can be defended, then the fair play account of punishment follows straightforwardly.

My primary focus in this paper, however, has been with the first point. That is, I contend that my fair play account is more plausible, as a fair play account of punishment, than are standard versions of the view. As I have discussed, my account leads to the implication that the question of punishment's in-principle permissibility is distinct from the questions of its positive aim and of how to punish in particular cases; answers to these distinct questions will require appeal to distinct moral considerations. Rather than regarding this implication as regrettable, however, I suggest that fair play theorists should embrace it. As I have argued, doing so is not only defensible in its own right, but it also allows the fair play view to avoid a number of unappealing implications.⁴⁶

⁴⁶ I am very grateful to Antony Duff, Larry May, Victor Tadros, Christopher Heath Wellman, David Wood, and an anonymous reviewer at *Criminal Law and Philosophy* for their helpful comments on earlier drafts of this chapter. Also, I presented a previous draft of the chapter at the Washington University in St. Louis Political Theory Workshop, February 2010. I thank the faculty members and students who participated, especially Nate Adams (my discussant), David Bauman, Jeff Brown, Adrienne Davis, Julia Driver, Chad Flanders, Clarissa Hayward, Frank Lovett, Ian MacMullen, David Speetzen, Ron Watson, and Carl Wellman, for their thoughtful questions and suggestions during and after the session.

CHAPTER 2

Deterrent punishment and respect for persons

In the previous chapter, I grounded punishment's in-principle moral permissibility in certain considerations of fairness shared by members of a political community. Even if punishment is permissible in principle, however, it may be unjustified all things considered if there is no legitimate, compelling reason to have such a practice. In my view, the reason we should want an institution of punishment is that it helps to protect community members' security and well-being by reducing criminal activity. In this chapter, I defend deterrence as the central aim of punishment.

Deterrence-based accounts of punishment have been criticized frequently because they are unable to rule out occasionally punishing innocent citizens, or disproportionately punishing guilty ones, if doing so would yield net deterrent benefits.¹ In response to these sorts of objections, some theorists have argued that although considerations of deterrence cannot ground a complete justification of punishment, they may nevertheless shoulder some of the justificatory burden. Perhaps most notably, H. L. A. Hart contended that consequentialist considerations such as crime prevention represent the central aim of

¹ See, e.g., David Boonin, *The Problem of Punishment* (New York City: Cambridge University Press, 2008), pp. 41-52; Deirdre Golash, *The Case Against Punishment: Retribution, Crime Prevention, and the Law* (New York City and London: New York University Press, 2005), pp. 43-44; John Braithwaite and Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* (Oxford, U.K.: Clarendon Press, 1990), p. 46; H. J. McCloskey, "Utilitarian and Retributive Punishment," *Journal of Philosophy* 64:3 (1967): 91-110, on pp. 93-102, and "An Examination of Restricted Utilitarianism," *The Philosophical Review* 66:4 (1957): 466-85, on pp. 468-69; and W. D. Ross, *The Right and the Good* (Oxford, U.K.: Oxford University Press, 1930), p. 57.

punishment, but that particular impositions of punishment should be constrained by the familiar principles that only the criminally guilty should be punished, and only in proportion with the seriousness of their crimes.² Constrained by principles such as these, deterrence as an aim of punishment looks significantly more appealing.

Even with these constraints, however, deterrence as an aim of punishment has been subject to a further line of criticism. Here, the objection is not that in some cases considerations of deterrence might permit the punishment of law abiders, but rather that punishment aimed at deterrence fails to respect offenders as autonomous moral agents — or in Kantian terms, as ends in themselves. This challenge is particularly powerful. It does not merely charge that deterrent punishment might allow, in certain cases, the disrespectful treatment of offenders; if this were the charge, then perhaps constraints could be articulated, similar to the constraints against punishing the innocent, to rule out such treatment. The objection here, however, is that punishment aimed at deterrence *by its nature* fails to treat offenders with respect. If the charge is valid, then additional constraints won't help.³

This paper defends deterrence as an aim (in my view, the central aim) of punishment against this objection that deterrent punishment fails to respect offenders as moral persons. I examine three prominent ways in which this charge has been fleshed out.

² H. L. A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (New York City: Oxford University Press, 1968), pp. 8-13. See also Don E. Scheid, "Constructing a Theory of Punishment, Desert, and the Distribution of Punishments," *Canadian Journal of Law and Jurisprudence* 10 (1997): 441-506. For a somewhat different sort of disaggregation of the relevant questions, see Ross, *The Right and the Good*, pp. 61-64.

³ C.f., Jeffrie G. Murphy, "Marxism and Retribution," *Philosophy & Public Affairs* 2:3 (1973): 217-43, on p. 219.

First, some theorists, such as Jeffrie Murphy, have objected that punishing with the aim of deterrence uses the offender as a mere means to secure some social benefit, namely, crime reduction. The second and third versions of the objection have been developed thoroughly by R. A. Duff. A system of criminal law and punishment aimed at deterrence, Duff claims, offers reasons for compliance that are inappropriate to autonomous moral agents, and it implicitly excludes criminals from membership in the political community. Duff offers these as aspects of the same line of critique, but I argue below that they are in fact separate charges and thus merit distinct consideration. I contend that none of these objections ultimately succeeds. That is, none of them establishes that punishment aimed at deterring crime fails to demonstrate appropriate respect for persons. Specifically, a deterrent system of punishment — bounded by appropriate constraints on who may be punished and how severely — does not treat offenders as mere means to securing certain social goods, it does not offer inappropriate reasons for compliance with the law, and it does not implicitly exclude criminals from membership in the political community.

In section I, I examine and refute the objection that deterrent punishment uses offenders as mere means to securing the social goal of crime reduction. In section II, I take a closer look at Duff's account and contend that he actually offers two distinguishable versions of the respect-based objection. In sections III and IV, I examine each of these objections in turn, and I conclude that neither succeeds. Ultimately, deterrence is a permissible aim for a system of criminal punishment; that is, punishment aimed at deterring crime can be consistent with respect for moral persons.

I. Does deterrence use offenders as mere means?

One way to interpret the charge that deterrent punishment fails to respect offenders as persons is that such punishment appears to use offenders as mere means to deterring crime. Jeffrie Murphy, for instance, has written of deterrence that “a guilty man is, on this theory, being punished because of the instrumental value the action of punishment will have in the future. He is being used as a means to some future good — e.g., the deterrence of others.”⁴ Such punishment thus appears inconsistent with maintaining proper respect for offenders as autonomous moral agents.

Murphy’s characterization of the good being sought as “the deterrence of others” points to a sense in which we might think one form of deterrence can be especially problematic. That is, it might seem bad enough that punishment subjects offenders to hard treatment with the aim of promoting the social good of crime reduction. A critic might further point out, however, that one type of deterrence, general deterrence, seeks to achieve this social good by treating offenders in certain ways in order to affect others’ behaviors, to persuade others to comply with the law. Special deterrence may also seem troubling insofar as it subjects an offender to hard treatment to bring about the social good of crime reduction, but at least it treats the offender in this manner with the aim of affecting her own future behavior, of persuading her to comply with the law in the future, rather than treating her in this way to affect others’ behavior. Thus insofar as this objection is valid, it strikes particularly hard at general deterrence.

⁴ Ibid., on p. 219. For similar articulations of this critique, see, e.g., R. A. Duff, *Punishment, Communication, and Community* (Oxford, U.K.: Oxford University Press, 2001), pp. 13-14; and Boonin, *The Problem of Punishment*, pp. 60-61.

We might respond to this line of critique by pointing out that political communities often harm law abiders for the sake of promoting some greater good, as well. Construction of a new highway may be beneficial to the community generally, but it may harm those who live nearby (perhaps by generating noise pollution or diminishing their property values). Similarly, those with a communicable disease may be forced to endure certain restrictions of their liberties in the interest of protecting public health. If harming some for the greater benefit of others is permissible in cases such as these, and numerous others, then perhaps harming offenders to benefit the public by deterring crime is similarly permissible.

David Boonin rejects this line of response, however, because he believes it overlooks the distinction between intending harm and foreseeing harm. Boonin points out that cases such as those described above — the highway construction, or quarantining those with a communicable disease — “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.”⁵ He continues:

[T]he 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

⁵ Boonin, *The Problem of Punishment*, p. 62.

punishment distinctively difficult to justify in the first place.⁶

Thus for Boonin, a deterrent system of punishment is objectionable because it intentionally harms some to benefit others. The harm is the means by which the good is achieved, not merely a foreseeable consequence.

Given that the ultimate aim of a system of deterrent punishment is to reduce crime, however, I suggest that actual inflictions of punishment are not the means by which the system seeks to achieve this aim. Rather, the *threat* of punishment is intended to do the deterrent work.⁷ A deterrent system of punishment communicates a threat to everyone in the community: If you do these acts, you will be subject to punishment. Consider that if the threat of deterrent punishment were perfectly effective, no one would violate the community's laws, and thus no one would be punished. Actual instances of punishment, then, are best seen as cases where the deterrent threat failed.⁸ The inflictions of harm that constitute punishment are not the means by which the good of crime reduction is achieved; rather, the means by which deterrent systems of punishment aim to reduce crime is by issuance of a threat. Obviously, in the actual world deterrent systems of punishment are not perfectly effective. Individuals continue to commit crimes despite the existence of the deterrent threat. In these cases, such individuals are harmed in the ways

⁶ Ibid. See also Nathan Hanna, "Say What? A Critique of Expressive Retributivism," *Law and Philosophy* 27 (2008): 123-150, on pp. 124-28.

⁷ One might worry that the threat of punishment is itself a sort of coercive sanction, in that its aim is to change incentives so that, in effect, it restricts citizens' viable options. I consider this point more below.

⁸ S. I. Benn noted this point in his "An Approach to the Problems of Punishment," *Philosophy* 33:127 (1958): 325-41, on p. 330.

characteristic of punishment. But such individuals are foreseeably rather than intentionally harmed. Again, this is because the intention of a deterrent system of punishment is that everyone should take the threat seriously and avoid criminal behavior (and, in turn, punishment).

Boonin insists, however, that the intended-harm element is essential to our conception not only of deterrent punishment, but of punishment in general. He writes:

When the state punishes someone, ... 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.⁹

Boonin may be correct with respect to punishment whose central aim is retribution, or perhaps even special deterrence (although even on these accounts there would presumably be plausible considerations cautioning against lengthening or altering sentences once they had been issued). But his point is mistaken with respect to general deterrence. In a system of punishment aimed at general deterrence, sentences are not imposed to inflict suffering on the offender, but rather to maintain a credible threat to the public generally. Typically, of course, the more severe the sentence, the more the offender will suffer *and* the more credible the threat will be. But the concern, from the perspective of general deterrence, is not how much an offender suffers, but rather how effectively the general public is deterred from committing the given offense. In fact, if the

⁹ Boonin, *The Problem of Punishment*, p. 13.

credible threat could be maintained without harming any offenders, then this would be entirely acceptable based solely on considerations of general deterrence.¹⁰ Punishment aimed at general deterrence, then, is best characterized not as intentionally harming some to benefit others, but rather as intentionally threatening everyone, and then foreseeably harming those who nevertheless commit crimes.

Suppose, however, that I am wrong about this. Suppose that punishment aimed at deterrence is in fact best understood as intentionally harming some to benefit others. This is still not enough to establish that such punishment would violate the Kantian principle of respect for persons. It's widely recognized that this principle does not forbid treating others as means, but as *mere* means. We frequently treat others as means to our own or other people's ends, and we typically consider such treatment permissible. I ask a taxi driver to take me to my destination, our country sends soldiers to fight in a war to protect our interests, I ask a friend to lend me money. The taxi driver, the soldiers, and my friend are all treated as means to others' ends (mine, or in the case of the soldiers, the country's), but we do not find these cases objectionable as long as they are not treated *merely* as

¹⁰ This point, in fact, grounds a distinct objection commonly leveled against deterrent punishments generally: Insofar as the deterrent threat is what is crucial, the legal authority might be justified in some cases of merely pretending to punish offenders. This prospect is particularly troubling to those with the retributivist intuition that the guilty deserve to suffer. Advocates of deterrence as an aim of punishment might respond to this objection in various ways: They might contend that the public's likely discovery of the pretend punishment cases could undermine the general deterrent effect of the threat; or that whereas general deterrence constitutes one aim of punishment, there are others (retribution, reform, etc.) that rule out the possibility of pretending to punish. Whether these or other responses would be persuasive need not worry us here, because the question of whether general deterrence would permit pretend-punishing is a distinct one from the question of whether punishments aimed at general deterrence treat offenders as mere means. For present purposes, the relevance of the pretend-punishing objection is that it underscores that punishment-as-general-deterrence aims to reduce crime not by harming offenders but rather by issuing a threat.

means. The relevant question for this version of the respect-based objection, then, is whether deterrent punishment treats offenders merely as means to the social good of crime reduction.

There are good reasons to doubt that punishing for deterrence uses offenders merely as means to the end of crime reduction. First, note that insofar as the institution of punishment does yield a deterrent effect, those who commit crimes typically will have reaped benefits from the existence of this institution just as law abiders have done. Perpetrators of crime are also, like other community members, potential victims of crime. Thus insofar as the institution of punishment helps to deter crime, it protects the safety and security of everyone.

One might respond that an offender may still be treated merely as a means when she is harmed in the interest of securing this social good, even if the social good is also a good for the offender herself. If our legal system sanctioned the occasional punishment of innocent people for the purpose of achieving the beneficial deterrent effect, for instance, then these individuals would be used as mere means even if they themselves had benefited from the deterrent effects of the institution generally. Thus even if offenders as well as law abiders enjoy the general benefits of deterrent punishment, this fact by itself appears insufficient to assure that such punishment avoids using offenders as mere means.

Deterrent punishment with prohibitions on punishing the guilty is relevantly different, however, from a system of deterrent punishment (even an overall beneficial one) that allowed the punishment of the innocent. To punish law abiders would be to treat them in ways that were not responsive to choices they had actually made, and thus it

would fail to respect them as autonomous moral agents. Respectful treatment requires at least that we treat others according to what they have actually done (or failed to do); punishing those who have violated no criminal laws fails to meet this minimal standard of respect. Notice, though, that a deterrent system of punishment does not only offer to each community member the benefits that come from reduced crime. Deterrent punishment constrained by the retributivist principle against punishing the innocent also allows each individual to choose whether she will risk suffering the harms associated with punishment. Such a system offers everyone a choice: Comply with the law, or be subject to punishment. Thus unlike the innocent person who is punished to achieve the deterrent effect, the offender's punishment is a response to the choice she made to violate the law. Given that her punishment is a response to her own free choice, the fact that the aim of punishing her is to deter others from committing similar crimes (or her from committing similar crimes in the future) does not imply that she is treated *merely* as a means to this end. Hart expresses essentially this idea, as he describes the institution of punishment as “*offering* individuals including the criminal the protection of the laws on terms which are fair, because they not only consist of a framework of reciprocal rights and duties, but because within this framework each individual is given a *fair* opportunity to choose between keeping the law required for society's protection or paying the penalty.”¹¹

Still, even if an institution of deterrent punishment offers benefits to everyone, and even if it offers each community member equally a choice about whether to endure

¹¹ Hart, *Punishment and Responsibility*, pp. 22-23.

the threatened sanction, one might still object that this choice itself is coercive, that it employs the threat of harm to restrict citizens' viable options. Richard Burgh objects to Hart's account by offering what he considers an analogous case, in which terrorists take a group of people hostage and tell each that if he tries to escape, he will be beaten. The terrorists treat all of the hostages equally, and they stay true to their pledge only to beat those hostages who try to escape. "Simply because a hostage is given a fair opportunity to avoid being beaten," Burgh concludes, "it does not follow that his beating is just."¹² Even if the terrorists "were to inform the hostages that if they do as they are told they will receive positive benefits," beating those who tried to escape would be unjust.¹³ Burgh concludes that, analogously, deterrent punishment cannot be justified on grounds that it provides a choice either to comply with the law and reap benefits from others' compliance or to break the law and suffer punishment.

Contrary to Burgh's charge, however, there is a fairly straightforward difference between the choice offered by the institution of deterrent punishment and the choice offered by the terrorists. Given that the terrorists violate each hostage's liberty rights, the hostages' choice is either not to do that which they have a moral right to do (namely, leave) or to be beaten. So the terrorists use the prospect of force to persuade the hostages not to do what they have a right to do. A system of deterrent punishment, however, employs the prospect of force to persuade community members not to do the sort of acts

¹² Richard Burgh, "Do the Guilty Deserve Punishment?" *The Journal of Philosophy* 79:4 (1982): 193-210, on p. 199.

¹³ *Ibid.*, p. 200.

that they have a moral obligation not to do.¹⁴ Thus the relevant question is whether a system of punishment that provided significant benefits to community members generally, and that offered a choice *either* not to commit acts that one has moral obligations not to commit *or* to be harmed, is coercive in a way that renders it inconsistent with respect for moral persons. Given that such a system offers each community member benefits, treats each according to her own choices, and seeks to persuade citizens not to do that which they have a moral obligation not to do anyway, I suggest that such a system is consistent with respecting individuals, even those punished, as autonomous moral agents.

Kant himself provides support for the view that punishment, properly constrained by the retributivist principle, may aim at deterrence while nevertheless respecting the offender. A criminal, he writes, “must previously have been found *punishable* before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens.”¹⁵ Although Kant’s full view of punishment continues to be the subject of substantial debate, in this passage he suggests that deterrence is a permissible aim, which for him means that it does not use the individual as a mere means, as long as punishments are limited to those who are guilty of crimes.

I conclude that punishment aimed at deterrence does not use offenders as mere means, and thus that this version of the respect-based objection fails. Still, Kant’s respect

¹⁴ This is the case, at least, when the laws backed by deterrent punishment are justified. By contrast, unjust laws (e.g., laws allowing, or requiring, what is morally prohibited) backed by deterrent punishment would be analogous to Burgh’s terrorist example. As such, a system of deterrent punishment backing such laws would fail to respect those punished as moral persons.

¹⁵ Immanuel Kant, *The Metaphysics of Morals*, 6:331. In Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary J. Gregor (Cambridge, U.K.: Cambridge University Press, 1996), p. 473.

principle instructs us not only not to use others as mere means, but also to respect them as ends in themselves. Respecting people as ends may require more than merely not using them as mere means. Perhaps, then, there is a sense in which punishment aimed at deterrence nevertheless violates the respect principle. In the following three sections, I consider what I take to be the most thorough and compelling development of this sort of objection, by R. A. Duff.

II. Duff's critique of deterrence

Duff conceives of the criminal law as fundamentally a communicative enterprise. He argues that a system of punishment that aims to deter potential offenders is inappropriate for a liberal political community, committed to respecting its members *as* members of the community. Essentially, this is because deterrent punishment communicates in prudential rather than moral terms:

The law of [a liberal political] community, as its common law, must address its members in terms of the values it embodies — values to which they should, as members of the community, already be committed. It portrays criminal conduct as wrongful in terms of those values; and the reasons that citizens have to refrain from such conduct, the reasons to which the law refers and on which it depends, are precisely the moral reasons that make such conduct wrong. A purely deterrent law, however, addresses those whom it seeks to deter, not in terms of the communal values that it aims to protect, but simply in the brute language of self-interest. It thus addresses them, not as members of the normative community of citizens, but as threatening outsiders against whom the community must protect itself. It implicitly excludes them from membership

of the citizen community by no longer addressing them in terms of that community's values.¹⁶

Duff is concerned, commendably, that offenders should be treated as moral persons, and in fact as continuing members of the community, rather than merely as the “they” against whom “we,” the law-abiding community members, must protect ourselves. His concern is well founded — it is all too easy, and too common, to assume that the criminal act necessarily demonstrates a criminal, perhaps even irredeemably criminal, character. Duff urges us, however, always to regard the person guilty of a criminal offense as nevertheless one of us, a member of our community who may come to share (or recommit to) the moral values that the community endorses. Despite the significant virtues of Duff's account, however, I contend that his objection to punishments aimed at deterrence misses its mark. There is a real sense in which a system of punishment aimed at deterring crime (with appropriate constraints) can nevertheless demonstrate appropriate respect for criminal offenders, and thus avoid being objectionably exclusionary.

Note that Duff actually offers two critiques of systems of punishment aimed at deterrence — two ways in which such systems of punishment fail to treat individuals with appropriate respect as autonomous moral agents. First, deterrent punishment offers individuals the wrong sort of reasons to comply with the law. It offers merely prudential reasons to comply — i.e., to avoid incarceration, community service, etc. — rather than the appropriate moral reasons — i.e., that the prohibited acts are morally condemned by the community. Second, by offering merely prudential reasons, rather than making the

¹⁶ Duff, *Punishment, Communication, and Community*, pp. 78-79.

sort of moral appeal that is appropriate to members of a liberal political community, a deterrent system of punishment implicitly excludes those it addresses from membership in the community. It fails to respect them as fellow community members who, as members, share (or should share, and can come to share) the community's moral values. Punishing to deter is thus exclusionary, Duff believes, in that it reinforces the distinction between "we," the law-abiding citizens, and "they," the criminals, rather than treating offenders as continuing to be fellow members of our community.

Duff implies that the second critique follows from the first. That is, he indicates that a system of punishment aimed at deterrence excludes certain individuals from the political community *because* it offers them the wrong sort of reasons (i.e., prudential reasons) to comply with the law. In fact, however, these are distinct critiques. The charge that deterrent punishment is exclusionary rests on the notion that it treats offenders differently from law abiders. It perpetuates the distinction between "us" (the law abiders) and "them" (the criminals) and implicitly excludes "them" from the community in which "we," as law abiders, are still included as members. By contrast, the objection that punishment aimed at deterrence provides the wrong sort of reasons to comply with the law does not depend on its offering *different* reasons to offenders and to law abiders. Rather, a system of punishment might offer the same, inappropriate reasons for compliance to everyone. As such, it would not treat one group (offenders) as less of a part of the political community than another group (law abiders). It would not perpetuate the objectionable "we" and "they" distinction, because such a system would communicate

the same message, and offer the same reasons, to everyone.¹⁷ Thus whereas one objection contends that deterrent punishment offers the wrong sort of reasons, the other contends that it inappropriately offers different reasons to different members of the community.

I suggest, then, that these two critiques warrant distinct consideration. We should ask, first, if punishment aimed at deterrence communicates a different message to (or provides different reasons to, or in some other way excludes) criminals from the political community generally; and second, if the reasons such a system of punishment offers for complying with the law are themselves the wrong sort of reasons to offer fellow members of the political community. I consider each of these critiques in turn and contend that, ultimately, each fails. A system of punishment aimed at deterrence communicates the same message to everyone in the political community, thus it does not implicitly exclude anyone. Furthermore, a deterrent system of punishment is compatible with demonstrating appropriate respect to all members of the community as members who share (or should share, and can come to share) the community's fundamental moral values.

III. Is deterrent punishment exclusionary?

The first objection evident in Duff's account is that a system of punishment aimed solely at deterrence implicitly excludes offenders from their community. It treats

¹⁷ One might respond that such a system of punishment would then be exclusionary of everyone. But if everyone is excluded, then we must ask, excluded from what? Duff's point is that deterrent punishments exclude offenders from their community, but if all community members (law abider and offender alike) were excluded, then it is not clear what community would remain for those excluded to be excluded *from*. Thus central to the charge that deterrent punishment excludes certain community members is the claim that it treats some (the excluded) differently from others (the included).

offenders as the “they” against whom “we,” the law-abiding members of the community, must protect ourselves. Thus it fails to treat offenders with appropriate respect. Given Duff’s conception of the criminal law as a fundamentally communicative enterprise, the worry with deterrent punishment is that, insofar as it offers the offender only prudential reasons why she should not have committed, say, theft or tax evasion, it fails to communicate with her as (still) a member of the community. A more appropriate message to a community member would appeal to the moral reasons that her act was wrong, namely, that such acts violate important moral values that the community shares (and thus that she, as a member of the community, should also share).

The thrust of the “exclusion” objection to deterrent punishment, then, is that once a member of the community commits a crime, the criminal law stops talking to her as it talks to law abiders, to whom it offers appropriate, moral reasons not to violate the law; instead, it begins to talk with her in the language of mere prudence, as though this is the only language she is capable of understanding. As such, it inappropriately excludes her from membership in her community.

One might understandably be tempted to respond here that the offender, in committing her crime, essentially excludes *herself* from membership in the community — or at least, that she demonstrates that she does not share the community’s moral values. If so, then it may seem appropriate for a system of punishment to communicate to her solely in the language of prudence rather than in the language of the community’s moral values. Duff rejects this argument, however, for several reasons, the most persuasive of which is that it is empirically dubious. Often, criminal acts are not evidence that

offenders have no regard for the community's moral values, but rather "that their regard is not wholehearted, or consistent, or always sufficient to overcome the temptations of self-interest. They — or rather *we*, since these comments surely apply to many of us — are not wholly deaf to the law's moral appeal, though we do not attend to it carefully or consistently enough"¹⁸ Duff is right to caution against assuming that an individual's criminal act is evidence of a complete rejection, or lack of regard, for the community's values.

There is a more fundamental problem, however, with the claim that deterrent punishment somehow communicates to offenders differently from law abiders, and thus excludes offenders from the political community. The message communicated by a system of punishment aimed at deterrence essentially takes the form of a threat: If you commit some criminal act, then you will be liable to have some form of suffering inflicted on you. It is important to consider, however, to whom this message is communicated. For deterrent punishment to be exclusionary, to create the sort of "we-they" dichotomy that concerns Duff, it would have to be the case that a system of deterrent punishment communicates one (prudential) message only to criminal offenders, and that law abiders, by contrast, receive another (moral) appeal that is appropriate to members of the political community.

But this is doubly wrong. First, a system of deterrent punishment communicates its prudential message, its threat, to everyone. For those who have not committed a crime,

¹⁸ Duff, *Punishment, Communication, and Community*, p. 84.

the prospect of punishment offers reasons not to do so (i.e., it acts as a general deterrent). For those who do commit crimes, their punishments — or more specifically, the unpleasant prospect of another term of punishment in the future — provide reasons not to recidivate (i.e., they act as a special deterrent). From the perspective of deterrent punishment, then, everyone is a potential offender (or reoffender), and such a system of punishment communicates the same prudential message to everyone. Therefore, second, if law abiders do receive the moral appeal that Duff believes is appropriate to members of the political community, the source of this appeal is not the system of deterrent punishment. Rather, the moral appeal must come from somewhere else, such as perhaps the criminal laws themselves. But if it is the criminal laws that communicate the moral message, that declare certain actions to be morally condemned by the community, the intended audience of this communication is everyone, law abider and offender alike. Thus it is not the case that, in receiving the prudential message of a deterrent system of punishment, offenders are treated differently from others in the community, who are exclusive recipients of the moral message. It appears that deterrent punishment is not essentially exclusionary in the way Duff indicates.

There are other ways, of course, in which existing penal practices tend to exclude offenders from the community. Imprisonment, by its nature, removes offenders physically from the larger community. Beyond this, prisoners are typically excluded from

participation in the political process, most obviously by being denied the vote.¹⁹ Also, offenders are excluded from access to basic financial services (bank accounts, credit, insurance), not only during their incarceration but often, in practice, even after their release.²⁰ These and other forms of exclusion should be troubling to members of a liberal political community who are concerned to treat individuals, even offenders, with respect as autonomous moral agents. But notice that such forms of exclusion are not distinctively characteristic of systems of punishment aimed at deterrence (and constrained in the ways suggested earlier). Because punishment involves the restriction of offenders' liberties in ways that law abiders' liberties are not restricted, issues of exclusion will always arise. But such issues are not distinctive of systems of punishment aimed at deterrence. Rather than communicating differently to offenders and law abiders, and thus perpetuating the "we-they" distinction that concerns Duff, systems of punishment aimed at deterrence regard everyone equally as potential offenders, and thus they communicate the same message, namely, if you commit a crime, then you will be liable to be harmed. I conclude, then, that punishment aimed at deterrence is not exclusionary as Duff charges.

¹⁹ In the United States, only Maine and Vermont allow incarcerated felons to vote. A number of states go further than this, imposing a lifetime ban on voting for anyone with a felony conviction, even those who have served their sentences.

²⁰ See Gaynor Pengelly, "Give prisoners bank accounts," *This is Money*, 25 October 2010, online at http://www.thisismoney.co.uk/savings-and-banking/article.html?in_article_id=517136&in_page_id=7&position=moretopstories (accessed 25 October 2010). For a fuller discussion of types of exclusion, see Duff, *Punishment, Communication, and Community*, pp. 75-77.

IV. Does deterrent punishment offer the wrong sort of reasons for compliance?

Given that deterrent punishment communicates the same reasons to everyone, the question then becomes whether these reasons are appropriate. Duff contends that they are not. He writes, “The criminal law of a liberal polity, and the criminal process of trial and conviction to which offenders are subjected, are communicative enterprises that address the citizens, as rational moral agents, in the normative language of the community’s values.”²¹ And the institution of punishment, a constitutive element of the institution of criminal law generally, must similarly communicate in moral rather than prudential terms. A system of punishment aimed at deterring criminals, however, aims to secure general compliance with the law by means of a threat, rather than by moral appeal. Thus Hegel famously objected: “To base a justification of punishment on threat is to liken it to the act of a man who lifts his stick to a dog. It is to treat a man like a dog instead of with the freedom and respect due to him as a man.”²² Like Hegel, Duff worries that the prudential terms in which deterrent punishment communicates with community members, the prudential reasons it gives them to comply with the community’s laws, are not the sort of reasons that are appropriate to offer to autonomous members of a liberal political community, who endorse (or could come to endorse) the community’s moral values.

I offer a couple of responses to this worry. First, punishment may communicate a prudential message to community members without communicating a *solely* prudential

²¹ Duff, *Punishment, Communication, and Community*, p. 80.

²² G. W. F. Hegel, *The Philosophy of Right*, trans. T. Knox (Oxford, U.K.: Oxford University Press, 1942), p. 246.

message. In my view, the good of punishment, the reason we should want such an institution, is that it plays a key role in ensuring the well-being of community members. Thus the proper aim of punishment is to prevent or reduce crimes by offering potential wrongdoers reasons not to offend. Punishment may serve this aim by supplying potential offenders with prudential reasons not to offend (reasons such as the desire to avoid the harms characteristic of incarceration, etc.), but it may also provide moral reasons. As is commonly recognized, punishment involves not only what Joel Feinberg called a “hard treatment” aspect but also an expressive aspect — punishment expresses the community’s condemnation of the offender for her criminal act.²³ Even before the commission of a crime, however, the threat of punishment also expresses the community’s condemnation not of a particular offender but rather of the offense itself. If a potential offender receives and accepts this message of condemnation, it may play a role in persuading her not to do what she otherwise would have done. If so, then even if the fear of punitive suffering also played a role in dissuading her, I suggest that she is treated with the respect due to her as a moral person. Thus even if a system of punishment’s central aim is to provide prudential reasons for compliance, this does not preclude its also providing moral reasons.

Second, even if a system of punishment did provide *solely* prudential reasons to comply with the law, this doesn’t show that the criminal legal system more generally fails to communicate with community members as moral persons. I agree with Duff that the

²³ Joel Feinberg, “The Expressive Function of Punishment,” in *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: Princeton University Press, 1970), pp. 95-118.

criminal law should appeal to citizens as moral agents who share (or should share, and can come to share) the community's values. But punishment is only one aspect of the criminal legal system. Suppose we grant, then, that a community through its criminal statutes declares certain acts to be wrong and makes a moral appeal to community members to comply, whereas trials and convictions communicate a message of deserved moral censure to the wrongdoer, and they urge the wrongdoer "to understand and accept the censure as justified . . ." ²⁴ Why, then, must punishment also make a moral appeal? Why is it inappropriate for the institution of punishment to communicate a solely prudential message?

First, one might argue that the criminal legal system must be univocal in the message it communicates to community members, and that this message must be a moral rather than a prudential one. Thus the institution of punishment, as one element of the criminal law generally, must communicate a moral message. It's not clear why this should be so, however. We can grant, with Duff, that the criminal legal institution should communicate a moral message to community members while still (a) recognizing that distinct elements of the institution can communicate different messages, (b) maintaining that the criminal statutes themselves, and perhaps the process of trial and conviction, sufficiently communicate the moral message, and thus (c) denying that punishment must communicate this same message. Notice, too, that if the entire criminal legal system must be univocal in its moral message, much more than deterrent punishment would be

²⁴ Duff, *Punishment, Communication, and Community*, p. 80.

prohibited. The practice of plea bargaining, for one, would appear unjustifiable if prudential appeals are inappropriate in criminal law. More reasonable, I suggest, is to claim that the criminal law should address community members in moral terms, and in fact that the moral message should be central, but that as long as this moral message is present, prudential appeals also have an appropriate role.

A second possible response is that whereas the criminal law need not, in principle, communicate only a moral message, the prudential message of deterrent punishment is inappropriate in practice because it tends to drown out the moral message. That is, perhaps the threat of punishment is so powerful that it tends to focus community members' attention on the prudential reasons not to commit crimes and cause them to lose sight of the moral appeal. Andrew von Hirsch, who conceives of punishment as offering prudential reasons to supplement the (sometimes insufficiently motivating) moral reasons supplied by the criminal law, advocates a "decremental strategy" according to which prescribed sentences would be reduced gradually to levels at which the prudential reasons they offered would not drown out the moral reasons for compliance with the laws.²⁵ Duff is skeptical of such a strategy, however, as he believes that sentences mild enough so as not to overwhelm the moral message with the prudential threat would be too mild to achieve much deterrent effect at all, whereas by contrast,

²⁵ Andrew von Hirsch, *Censure and Sanctions* (Oxford, U.K.: Oxford University Press, 1996), chapter 5.

sentences sufficiently severe to provide any genuine deterrent effect would replace, rather than merely supplement, the moral appeal.²⁶

I suggest that this worry, that the prudential appeal of deterrent punishment may drown out the moral message of the criminal law generally, inaccurately depicts the relationship of the moral and prudential appeals. Rather than accepting that a stronger prudential message will tend to weaken the moral message comparatively, why not acknowledge that the prudential threat actually can reinforce the moral appeal? Granted, an institution of punishment aimed at deterrence provides prudential reasons to comply with the community's laws. But the existence of such an institution also invites us to consider, or remind ourselves, why our community believes that these laws, and the interests they protect, are of sufficient moral weight that we are willing to invoke the threat of hard treatment to help ensure that they are not violated. Rather than drowning out the moral message of the criminal law, as Duff fears — the message that certain acts are prohibited because society regards them as significant moral violations — deterrent punishment can reinforce this message, as it underscores that protecting community members from such violations is sufficiently important to warrant the infliction of harm as a response.

The prudential message of deterrent punishment, therefore, is compatible with the criminal law's nevertheless communicating a moral appeal to community members, and thus with respecting them as autonomous moral agents. On one hand, a system of

²⁶ Duff, *Punishment, Communication, and Community*, p. 88.

punishment aimed at deterrence may nevertheless provide moral as well as prudential reasons for compliance. On the other hand, even if punishment itself provides only prudential reasons, the criminal legal system need not be univocal. As long as the moral message is communicated prominently (by the laws themselves and the process of trial and conviction), respect for persons does not require that the institution of punishment communicate in moral terms. I conclude, then, that this third articulation of the respect-based objection fails.

V. Conclusion

In closing, it's worth emphasizing again the scope of the defense of deterrence that I have offered here. Specifically, I have not aimed to defend deterrence as sufficient to ground a complete justification of punishment. Rather, my focus has been on deterrence as the aim of punishment, constrained by certain considerations such as the retributivist principles against punishing the innocent or excessively punishing the guilty. Critics of deterrence, such as Murphy and Duff, claim that even as one element of this sort of hybrid account of punishment, deterrence is objectionable because it fails to treat individuals with appropriate respect as autonomous moral agents. I have contended, however, that on what I take to be the three most plausible articulations of this critique, it nonetheless fails. Constrained in certain ways, then, I conclude that deterrence is a permissible aim of punishment.²⁷

²⁷ I am grateful to Larry May, Christopher Heath Wellman, and David Wood for their helpful comments on earlier drafts of this paper.

CHAPTER 3

Retributivism as a constraint on punishment

I claimed in the chapter 2 that deterrence is a legitimate aim of punishment if we accept certain constraints to ensure that particular impositions of punishment are consistent with respect for moral persons. In this chapter and chapter 4, I focus on what sort of constraints are warranted. An ostensible virtue of retributivism is that it entails intuitively compelling limitations on punishments, namely, that only the guilty may be punished and that punishments should be no more severe than is deserved as a response to the crime (viz., punishment should fit the crime). Proponents of such a constraint worry that punishment grounded purely in consequentialist considerations such as deterrence may not have the resources to rule out punishment of the innocent or excessive punishment of the guilty, insofar as there could be cases in which such punishment would promote the best overall consequences. In Kantian terms, the retributivist constraints are intended to ensure that those punished are treated with respect as moral persons. These constraints are viewed by many legal theorists and practitioners not only as necessary to ensure that those punished are appropriately respected, but also as sufficient.

My focus in what follows is on the second retributivist constraint, that punishment should be no more severe than is deserved. Critics of retributivism point to its notorious inability to make sense of the notion of desert, or of a punishment's fitting a crime. Many see it as a fatal flaw of retributivism that it doesn't have the resources to provide guidance

about what punishments are deserved for various crimes. Russ Shafer-Landau, for instance, writes:

I do not believe that we can make sense of commensurating punishment with moral desert. If we can't, then the commensurability thesis [that sentencing guidelines are morally justified if and only if they assign punishments commensurate with moral desert] is false. And if the commensurability thesis is false, so too is retributivism.¹

In this chapter, I argue that the retributivist principle is indeed insufficient to ensure adequate constraints on punishment — in particular, it is unable to provide sufficient protection against overly harsh punishment. But this is not, at least not primarily, for the reason that critics such as Shafer-Landau suggest. On the contrary, considerations of moral desert may ground genuinely useful, albeit imperfect, guidance in sentencing determinations. The more serious problem with retributivism is that, even if it could provide definitive guidance regarding what punishment is morally deserved for a given crime, in some cases this punishment will nevertheless strike many as too severe. What treatment an offender morally deserves all things considered will often depend on more factors than those on which retributivism focuses: the seriousness of a person's crime and her degree of responsibility for it. Retributivism itself provides no basis for taking considerations other than these into account in determining how a political authority should punish. Thus retributivism is insufficient as a constraint on punishment.

¹ Russ Shafer-Landau, "Retributivism and Desert," *Pacific Philosophical Quarterly* 81 (2000): 189-214, on p. 190.

Insofar as we are concerned to treat even offenders as ends in themselves, some additional constraint(s) will be warranted.

In section I, I flesh out the retributivist constraint and examine the objection often raised against retributivism that there is no adequate way to determine what punishment is morally deserved for a given crime. I focus on Shafer-Landau's particularly thorough development of the objection, which, using an argument by elimination, concludes that there is no fact of the matter about what punishment an offender deserves for a crime. Thus his account includes not only the epistemological claim that we cannot know what punishment a given crime morally deserves, but also the metaethical claim that there is no fact of the matter to be known. In section II, after briefly taking issue with the metaethical claim, I turn my focus to the epistemological worry: Given the inevitable, apparently irresolvable disagreements about what punishments are morally deserved in various cases, is the retributivist injunction on punishments that exceed what is deserved essentially useless as a practical constraint on sentencing determinations? In response, I contend that we may (at least if we reject nihilism about moral desert) believe this retributivist constraint has some value insofar as (a) it may provide some guidance as a general moral principle from which particular moral judgments may be, albeit imperfectly, inferred, and (b) it ensures that desert, rather than various consequentialist considerations, is the target in determining sentences. The deeper problem with the retributivist constraint, however, lies in retributivism's overly narrow criteria in determining what response a wrongdoer morally deserves. In section III, I contend that to treat offenders with respect as moral persons, it is not sufficient that we ensure that their

punishments fit their crimes. Respect requires some additional, nonretributivist constraint(s) on punishment. Defending such a constraint is my project in chapter 4.

I. The retributivist constraint, and the standard objection

Theorists who endorse the retributivist constraint that punishments should be only to the degree morally deserved (no excessive punishment)² see it as a valuable check on a political authority's ability to impose overly harsh sentences to further some consequentialist goal. Two features of this constraint are worth noting. First, as typically endorsed — viz., punishments may be *only* as severe as is morally deserved — this constraint is sometimes referred to as a *negative* retributivist principle, and it is contrasted with the *positive* retributivist view that punishments should be no more *or less* than is morally deserved.³ Negative retributivism (sometimes called minimalism) thus differs from positive retributivism in that the former, but not the latter, could permit punishment less severe than is believed to be deserved. For instance, J. Angelo Corlett endorses negative retributivism and suggests as reasons for decreased punishment of the guilty, “plea bargaining for the sake of securing stronger punishments for greater offenders who deserve it, or simply not punishing minor offenses so that limited resources can be

² See, e.g., W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930), pp. 62-63; James Sterba, “Retributive Justice,” *Political Theory* 3 (August 1977): 349-362; David Wood, “Reductivism, Retributivism, and the Civil Detention of Dangerous Offenders,” *Utilitas* 9:1 (1997): 131-146, esp. p. 132; and J. Angelo Corlett, “Making Sense of Retributivism,” *Philosophy* 76:295 (Jan. 2001): 77-110.

³ C.f., J. L. Mackie, “Morality and the Retributive Emotions,” *Criminal Justice Ethics* (1982): 3-10; R. A. Duff and David Garland, *A Reader on Punishment* (Oxford: Oxford University Press, 1994), pp. 7-8; R. A. Duff, “Legal Punishment,” *Stanford Encyclopedia of Philosophy* (2008), at <http://plato.stanford.edu/entries/legal-punishment/> (accessed Sept. 8, 2008); and John Cottingham, “Varieties of Retribution,” *The Philosophical Quarterly* 29 (1979): 238-46, at pp. 240-42.

focused on more important wrongdoings.”⁴ Much of the popularity of the negative retributivist constraint is that this constraint prohibits excessive punishment but also allows for the intuitively attractive possibility that the mitigation of punishment may, in some cases, be justified. The negative retributivist constraint is thus widely endorsed, both by those (such as Corlett) who offer purely retributivist accounts of punishment and also those (such as Ross and Hart) who endorse a consequentialist aim for punishment as well as nonconsequentialist constraints.⁵

Second, the requirement that offenders should be punished only to the extent they deserve is best understood as a claim about moral rather than legal desert. It’s true that, in speaking of what punishment a criminal deserves, one might refer solely to the legally deserved punishment, i.e., the punishment (or range of punishments) authorized by sentencing guidelines. But this is no help if what we trying to determine is whether the legally sanctioned sentences are themselves justified. The existence in some society of a law according to which the offense of jaywalking is punishable by death would not convince most of us that jaywalkers in such a society therefore deserve to die; rather, most of us would feel that the law is unjust. What the retributivist needs, then, is an account of moral desert, not merely legal desert. As Ted Honderich writes, “We get no conceivable moral justification for punishment, no obligation or permission to punish,

⁴ J. Angelo Corlett, “Making Sense of Retributivism,” *Philosophy* 76:295 (Jan. 2001): 77-110, on p. 78, fn. 7.

⁵ Ross, *The Right and the Good*, pp. 56-64; H. L. A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (Oxford: Oxford University Press, 1968), pp. 1-13.

from the fact that an act was against a law, no matter how wrong or useless or disastrous the law, no matter the worth of the whole body of law of which it is part.”⁶

A common objection to the claim that punishment should be only as severe as is morally deserved, and in fact to retributivism in general, is that there is no apparent way to determine what punishment a given crime morally deserves. Those who would try to link an offense directly to some punishment it deserves will face the problem of explaining in what respect, in virtue of what, the punishment is deserved. Those who would instead attempt first to rank crimes from least serious to most serious, and then to map this ordinal ranking onto sentences that similarly run from least severe to most severe, do not avoid the problem. Such accounts would need not only an explanation of why one crime merits more severe punishment than another, but also of how to map the ordinal ranking of crimes onto the penalty schedule. For a retributivist, these questions will all need to be answered in terms of moral desert, but again, it’s unclear *how* desert should be understood to link crimes and punishments.⁷ Shafer-Landau describes the problem this way:

⁶ Ted Honderich, *Punishment: The Supposed Justifications Revisited* (London: Pluto Press, 2006), p. 23.

⁷ Among those who have cited concerns about retributivists’ inability to settle questions of moral desert, either directly or as part of a proportional mapping of punishments onto crimes, are S.I. Benn, “An Approach to the Problem of Punishment,” *Philosophy* 33:127 (Oct. 1958): 335-37; Edmund L. Pincoffs, “Are Questions of Desert Decidable?” in J.B. Cederblom and William L. Blizek, eds., *Justice and Punishment* (Cambridge, Mass.: Ballinger Publishing Co., 1977), pp. 75-88, esp. pp. 84-86; Hugo Adam Bedau, “Retribution and the Theory of Punishment,” *The Journal of Philosophy* 75:11 (Nov. 1978): 601-20, on pp. 610-15; Andrew Ashworth, “Criminal Justice and Deserved Sentences,” *Criminal Law Review* (1989): 340-55, on pp. 344-46; Michael Tonry, “Proportionality, Parsimony, and Interchangeability of Punishments,” in R. A. Duff and David Garland, eds., *A Reader on Punishment* (Oxford: Oxford University Press, 1994), pp. 145-51; and Honderich, *Punishment: The Supposed Justifications Revisited*, pp. 36-41.

If the commensurability thesis is true, and if legal punishment can be morally justified, then there must be some sanction or range of sanctions that a criminal morally deserves for his conduct. But consider some standard cases and see the difficulty for yourself: how much suffering is morally deserved for one who impersonates an officer, or counterfeits currency, or hijacks an airplane, or batters a child? ... [There] does not appear to be any way to know whether the impersonator morally deserves eighty, eight hundred or eight thousand days behind bars, or even whether some amount of jail time is the appropriate kind of punishment to impose in the first place. We can allow for some indeterminacy in the sentencing correlations, but at some point we must ask whether moral desert is giving us any guidance at all.⁸

. Although the critique of retributivism for its inability to settle questions of moral desert is quite common, Shafer-Landau's essay is distinctive, as far as I can tell, in that it not only cites this difficulty, but thoroughly illustrates how the difficulty arises for the various versions of retributivism. He sets about an argument by elimination, as he considers and rejects a number of retributivist attempts to answer the question of what punishment a crime morally deserves. Here I consider five of the more prominent of these (which, I take it, are sufficient to demonstrate the general problem for retributivists in determining moral desert): (1) punishment should be equal in kind to the crime; or (2) it should inflict equal suffering; or (3) it should be a product of the seriousness of the crime multiplied by the offender's degree of responsibility; (4) it should correct the unfair advantage gained by the perpetrator; or (5) it should communicate public condemnation of the crime.

⁸ Ibid.

First, there is the famous law of retribution, *lex talionis*, often expressed as “an eye for an eye, a tooth for a tooth,” which holds that a criminal should receive treatment equal in kind to that which she inflicted on her victim. Kant famously endorses *lex talionis* in his *Doctrine of Right*, where, discussing the appropriate amount and kind of punishment for a given crime, he writes: “whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself. If you insult him, you insult yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself.”⁹ As Shafer-Landau points out, problems with this sort of strict interpretation of *lex talionis* quickly become apparent.¹⁰ For many offenses, an in-kind response is impossible, either because there is no clear victim (e.g., many cases of reckless endangerment); the harm is spread over numerous victims, each of whom is only negligibly harmed (e.g., tax fraud, or vandalism of public property); or the perpetrator is in a relevantly different situation from her victim (e.g., kidnapping by a childless person). Thus this strict interpretation of the law of retribution, according to which crimes deserve punishment that is equal in kind, seems implausible in many cases.

Attempts to modify *lex talionis* to address these concerns do not fare much better. For instance, some have argued that the principle is best interpreted not as endorsing an in-kind response, but rather punishment that inflicts an equal amount of suffering on the

⁹ Kant, *Metaphysics of Morals* 6:332. In Mary J. Gregor, trans. and ed., *The Cambridge Edition of the Works of Immanuel Kant, Practical Philosophy* (Cambridge, U.K.: Cambridge University Press, 1996). All citations to the *Metaphysics of Morals* are to this edition.

¹⁰ Shafer-Landau, “Retributivism and Desert,” p. 193.

criminal as her crime inflicted on the victim(s).¹¹ But as Shafer-Landau indicates, this version of the law of retribution encounters similar problems to the equal-in-kind version.¹² How much suffering is deserved, for instance, for reckless endangerment, or for tax fraud? Further, how do we make interpersonal comparisons of suffering? Also, for both the above interpretations of *lex talionis*, in which desert is wholly determined by the harm done to the victim, there seems no room for consideration of *mens rea* (criminal intent), although intent is typically held to be directly relevant to sentencing decisions.

A third sort of retributivist answer, offered by Robert Nozick, attempts to deal with the *mens rea* concern by explicitly building considerations of responsibility into his account of moral desert.¹³ Essentially, on Nozick's account, we determine desert by multiplying an offender's degree of responsibility for an offense, r , by the wrongness of the offense, H . On this formula, the value of r may range from 1 (full responsibility) to 0 (no responsibility), and the value of H "is a measure of the wrongness or harm, done or intended, of the act."¹⁴ Thus an offender deserves punishment proportionate to H when he is fully responsible for his crime ($1 \times H = H$), he deserves no punishment when he is not

¹¹ See, e.g., Jeffrey Reiman, "Justice, Civilization and the Death Penalty: Answering van den Haag," *Philosophy & Public Affairs* 14 (1985): 115-48, esp. p. 125.

¹² Shafer-Landau, "Retributivism and Desert, pp. 193-94.

¹³ Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981), pp. 363-97.

¹⁴ *Ibid.*, p. 363.

at all responsible ($0 \times H = 0$), and “otherwise punishment is discounted by (because multiplied by) the person’s intermediate degree of responsibility.”¹⁵

Nozick’s account improves on the two previous accounts insofar as it acknowledges the role of responsibility in determining moral desert. Still, this account still doesn’t provide definitive guidance in sentencing. After all, it’s unclear how we are to quantify the person’s responsibility, r , in any of the intermediate stages between full and no responsibility. Similarly, Nozick provides no adequate account of how to determine degrees of wrongness, H . He indicates only that the H -value for an offense should be whichever is greater, “the amount of disutility the victim reasonably could have been expected to undergo, [or] the amount of disutility the perpetrator would (reasonably be expected to?) undergo from that same act.”¹⁶ Even if we find this claim compelling in principle, it offers no guidance in how to measure disutility, or for that matter, what would count as a reasonable expectation. So it is unclear on Nozick’s view how to determine values for r or H , and because r and H are his determinants of moral desert, this account cannot avoid the objection that it offers insufficient guidance in determining moral desert.

A fourth retributivist account, which I discussed in chapter 1, is the fair play view.¹⁷ According to the standard articulation of this view, the wrong to be redressed by

¹⁵ Ibid.

¹⁶ Ibid., p. 365.

¹⁷ Note that the nonstandard version of the fair play view that I defended in chapter 1 was not a retributivist account. Among the most prominent articulations of this view as a retributivist justification of punishment are Herbert Morris, “Persons and Punishment,” *Monist* 52 (1968): 475-501, and Jeffrie Murphy,

punishment is a wrong against society generally. Essentially, the idea is that we all benefit from our society's laws, which protect against interferences with our liberties; but along with this benefit comes the corresponding responsibility to refrain from breaking these laws. A criminal, like other members of society, benefits from the general obedience to laws, but she fails to reciprocate by obeying the laws herself. By failing to restrain herself appropriately, she gains an additional degree of liberty, an unfair advantage over the rest of society (in a sense, she becomes a free rider), and the justification of punishment is that it corrects this unfair advantage by inflicting harm on the offender proportionate to the benefit she gained by committing her crime.

This version of retributivism aims to avoid the problems of the earlier versions with determining punishment deserved where there is no clear victim; on the reciprocity view, punishment is a response not to a specific harm done to some specific victim(s), but rather to an unfair advantage taken against society. Reciprocity theories still face two significant problems, however: First, as has been commonly noted, it is counterintuitive to think of the wrong perpetrated by, e.g., a rapist as a sort of free-riding wrong against society in general, rather than an egregious wrong perpetrated against the victim. Second, even if we grant that the relevant harm to be addressed by punishment is the additional liberty unfairly enjoyed by the offender, how are we to determine the deserved punishment, which on this account would be the punishment sufficient to offset, or nullify, the criminal's unfair advantage? What degree of unfair advantage over society,

Retribution, Justice, and Therapy (Boston: D. Reidel Publishing Co., 1978), esp. pp. 82-115 and 223-49. For a useful critical discussion of reciprocity theories, see M. Margaret Falls, "Retribution, Reciprocity, and Respect for Persons," *Law and Philosophy* 6:1 (1987): 25-51, on pp. 27-38.

how much additional liberty, is gained by, e.g., a child molester? And what type and degree of punishment would be sufficient to offset the unfair advantage?¹⁸ Ultimately, it's not clear that the fair play view gets us much closer to an answer to the moral desert question than do the other retributivist efforts.

A fifth, increasingly popular retributivist answer to the question of how much punishment is deserved is offered by the retributivist strain of communicative theories. Communicative theorists contend that punishment is justified insofar as it expresses a message of public moral condemnation, or censure. Typically, modern communicative theories incorporate a retributivist element: the condemnatory message is justified because it is morally deserved.¹⁹ In particular, Shafer-Landau focuses on a retributive-communicative view espoused by Jean Hampton, according to which the wrongness of crimes is that they demean their victims, or communicate a message of their victims' inferiority. Thus punishment is justified insofar as it sends a contrasting, nullifying

¹⁸ Perhaps the most sustained, if ultimately unsuccessful, attempt to answer these questions is offered by Michael Davis. See, e.g., "How to Make the Punishment Fit the Crime," *Ethics* 93 (July 1983): 726-52; and "Using the Market to Measure Deserved Punishment: A Final Defense," in Davis, ed., *To Make the Punishment Fit the Crime: Essays in the Theory of Criminal Justice* (Boulder, Colo.: Westview, 1992), pp. 234-53. For objections to Davis' account, see Shafer-Landau, "Retributivism and Desert," pp. 206-07; and David Boonin, *The Problem of Punishment* (Cambridge, U.K.: Cambridge University Press, 2008), pp. 129-35.

¹⁹ On other censure theories, the message communicated by punishment may be justified on nonretributivist grounds. Emile Durkheim offers an account on which the message communicated by punishment is justified insofar as it promotes social solidarity (Durkheim, *The Division of Labor in Society*, trans. G. Simpson (New York City: The Free Press, 1964)). An account by A.C. Ewing claims that punishment is justified as a valuable communication that morally educates the public (Ewing, "Punishment as Moral Agency: An Attempt to Reconcile the Retributive and the Utilitarian View," *Mind* 36:143 (July 1927): 292-305).

message to victim and perpetrator, a message that affirms their equal status.²⁰ Like the views discussed above, however, Hampton's retributivist-communicative theory faces difficulties in determining what manner and degree of punishment is deserved in cases with no clear victims. Also, even in cases with clear victims, it's not clear that the relevant feature of such crimes, the wrong that deserves to be punished, is that they demean their victims, or express their inferiority. Shafer-Landau points out that, e.g., thieves or embezzlers need not feel superiority over their victims, and in any case this doesn't seem to be what is centrally wrong about such crimes.²¹

Other retributivist-communicative accounts may fare better on some of these points. For example, a Kantian account offered by M. Margaret Falls contends that punishment is justified in that it respects wrongdoers by holding them morally accountable for their crimes, and that holding offenders accountable requires that sentences should "appropriately communicate the state's condemnation of the criminal's deed."²² Falls' account determines the proper communication of condemnation in terms of what would hold the offender appropriately accountable, rather than what would reaffirm the victim's equal status, thus it seems less susceptible to Shafer-Landau's worry about determining desert in cases with no clear victims. Nevertheless, Falls' account

²⁰ Jean Hampton, "A New Theory of Retribution," in C. Morris and R. Frey, eds., *Liability and Responsibility* (Cambridge, U.K.: Cambridge University Press, 1991), pp. 377-414, esp. pp. 400-05. Hampton believed her view captured what Hegel had in mind when he famously wrote that punishment "annuls the crime." She wrote, "Of course it can't annul the act itself, but *it can annul the false evidence seemingly provided by the wrongdoing of the relative worth of the victim and the wrongdoer*" (p. 403).

²¹ Shafer-Landau's discussion of Hampton's view is in "Retributivism and Desert," pp. 195-97.

²² Falls, "Retribution, Reciprocity, and Respect for Persons," p. 45.

ultimately fares no better than Hampton's in terms of offering definitive sentencing guidance. For Hampton, the unanswered question is what mode and degree of punishment properly nullifies the demeaning message sent by, e.g., the rapist or the armed robber. For Falls, the unanswered question is what mode and degree of punishment properly holds the rapist, the armed robber, etc., morally accountable for his act. Whichever the question, retributivist-communicative accounts appear unable to provide a clear, specific answer.

Shafer-Landau considers several other retributivist accounts but finds them all ultimately inadequate in providing concrete guidance regarding morally deserved sentences. Given the inability of these various retributivist accounts to provide genuine guidance regarding the question of sentencing, what mode and degree of punishment is morally deserved, Shafer-Landau concludes that we have good reason to endorse nihilism about moral desert, according to which "there is no fact of the matter about what sanction(s) a wrongdoer morally deserves for his offense."²³ If nihilism about moral desert is true, he contends, then the retributivist thesis that punishments should be commensurate with moral desert is false. In the following section, I first briefly suggest reasons to reject his nihilist thesis. Even if nihilism about moral desert is false, however, so that there is some fact of the matter about what punishments are deserved for what crimes, one might wonder how much this helps if we do not, perhaps cannot, know this truth about moral desert. Thus my main task in the following section is to consider

²³ Shafer-Landau, "Retributivism and Desert," p. 191.

whether this epistemological hurdle fatally undermines the practical use of retributivist considerations for our sentencing determinations.

II. Two ways the retributivist constraint may still be useful

Shafer-Landau concludes, based on the inability of retributivist accounts to give a satisfactory answer to the question of how much punishment is deserved for a given crime, that there is no answer to this question, because there is no fact of the matter about moral desert. It's worth noting, however, that our persistent uncertainty about moral desert is also consistent with the conclusion that there is a truth about moral desert (a truth that either exists objectively, i.e., moral realism, or has its grounding in the views, whether actual or somehow idealized, of human beings, i.e., moral constructivism).²⁴ My intent here is not to dive headlong into the interesting metaethical debate between those who believe there are moral truths and those who do not. Indeed, I'm not sure how such a debate could be ultimately settled, given that our current situation of moral uncertainty is compatible with nihilism as well as with either realism or constructivism. But it is worth pointing out that the claim that there are moral truths fits more neatly with our pretheoretical intuitions. When we say that, e.g., the murderer got the punishment he deserved, we certainly speak as if there is a fact of the matter about moral desert (notice that even those who may disagree about what is deserved in such a case will each tend to speak as if there is a fact of the matter about desert). Our intuitions tend to be particularly

²⁴ Ironically, Shafer-Landau himself provides a particularly compelling, sustained defense of moral realism, in his aptly titled *Moral Realism: A Defence* (Oxford: Oxford University Press, 2003).

strong when we move to either side of the spectrum, to cases in which we believe the punishment was *not* what was deserved, but rather was either clearly deficient (e.g., a corporate polluter that received a proverbial slap on the wrist²⁵) or clearly excessive (e.g., a group of students severely caned for littering²⁶). Appeals to intuition obviously won't disprove nihilism about moral desert, but my aim here is just to point out that, based on our current inability to settle the question of what punishment a given crime deserves, we need not conclude that there is no answer to this question. There may be some punishment or range of punishments that is deserved in each case, some truth about moral desert.

Even if there actually is in each case a correct answer to the question of what punishment is morally deserved for a crime, it nevertheless remains true that we have been unable to pin down with any certainty what those deserved punishments are. As a practical matter, then, a question persists as to whether the retributivist constraint on excessive (i.e., undeserved) punishments has any useful role to play in sentencing determinations. I suggest two ways in which this constraint can be seen as genuinely useful, albeit insufficient.

First, note that the difficulty Shafer-Landau and others point out is essentially one of deriving specific guidance, in particular instances, from some general moral principle(s). The general moral principle at issue is that one should not be punished in

²⁵ C.f., Fanny Carrier, "Top U.S. court slashes Exxon Valdez oil damages," *Agence France Presse*, 25 June 2008.

²⁶ C.f., Sulok Tawie, "Now, case of excessive caning by teacher," *New Straits Times* (Malaysia), 23 July 2007.

excess of what she deserves, but the difficulty comes in trying to determine what this general principle implies in specific cases. Understood this way, it becomes clear that the problem is not one unique to moral desert. (Consider, for instance, the difficulty in trying to determine what constitutes a just distribution of societal resources.) In fact, the problem will arise whenever we try to move from general moral ideals, principles, or obligations to moral judgments in particular cases.

W. D. Ross discusses this problem in his well-known account of *prima facie* duties.²⁷ Ross endorses a pluralistic normative ethical theory; on his account, we intuitively recognize a number of *prima facie* moral duties, some consequentialist but also several nonconsequentialist. On Ross' account, these are duties of fidelity, reparation, gratitude, justice, beneficence, self-improvement and nonmaleficence.²⁸ These various duties are *prima facie* in the sense that they may be overridden in certain cases, but they will nevertheless continue to be recognized as duties. So we have, for example, a *prima facie* duty to keep our promises (fidelity), but there may be certain circumstances in which, for instance, our *prima facie* duty of beneficence overrides the duty of fidelity. I promise to meet you for lunch, but as I drive along the lakeside drive to meet you, I notice a child drowning and am in a position to help. My duty to help the child overrides my duty to keep my promise to you, and so my all-things-considered obligation is to stop

²⁷ Ross, *The Right and the Good*, pp. 16-47.

²⁸ *Ibid.*, pp. 21-22.

and try to save the child (even though I still recognize my broken promise to you as a duty that I did not fulfill).

Ross' account is irreducibly pluralistic; any of his seven basic, *prima facie* duties may override any others in the right circumstances. The question arises, then, of how we are to draw guidance from these general obligations in particular cases; how do we weight them, and thus determine which one, or group of them, overrides which others? Ross recognizes this problem, as he writes that the correct choice in a particular case is neither self-evident nor a logical conclusion inferred from the more general *prima facie* duties. But despite the fact that there is no clear, systematic way to infer particular judgments from the general obligations, these general obligations can nevertheless provide some guidance. As Ross writes: "we are more likely to do our duty if we reflect to the best of our ability on the *prima facie* rightness or wrongness of various possible acts in virtue of the characteristics we perceive them to have, than if we act without reflection. With this greater likelihood we must be content."²⁹ Like Aristotle, Ross thus indicates that there is no clear moral decision procedure; rather, arriving at the morally correct decisions in particular cases is an exercise in practical reason.

My suggestion is that, similarly, although there may be no precise, straightforward inferential path from the general moral principle that offenders should be punished only to the extent of their moral desert to the correct, morally deserved sentence, the ideal of moral desert may nevertheless offer some guidance as judges

²⁹ Ibid., p. 32.

attempt, in exercises of practical reason, to arrive at appropriate sentencing determinations in particular cases.³⁰ If so, it may be that some of the various retributivist accounts of desert may prove useful after all, not because any one of them specifically implies a determinate deserved punishment in every given case, but rather because, insofar as some of these accounts reflect our intuitions about an aspect of moral desert, they may provide at least *some* guidance in determining desert in *some* cases. So, for instance, Shafer-Landau rightly points out that retributivist accounts according to which an offender deserves only as much suffering as she inflicted on her victim will be entirely unhelpful in determining sentences for crimes with no clear victims. But in cases in which there are clear victims (assault cases, for instance), the principle that an offender deserves no greater degree of suffering than he inflicted on his victim seems a relevant consideration in the exercise of practical reason to determine an appropriate sentence. Similarly, fair play accounts appear ill-suited to prescribe appropriate sentences for crimes that are centrally wrongs done to some particular victim(s), rather than cases of free-riding on society as a whole. But in other cases (e.g., tax evasion), the principle that an offender deserves no more punishment than is required to offset the unfair advantage she gained over other members of the community may be an appropriate consideration in determining a suitable sentence. Again, these considerations will not directly imply any

³⁰ Note that this proposal implies a presumption against practices such as mandatory sentencing guidelines, in favor of more flexibility for judges: Because the sentence that reflects the general ideal of moral desert will differ from circumstance to circumstance, our sentencing policies should allow for more fine-grained exercises of practical reason. This is, however, a presumptive consideration, rather than an absolute one, in favor of greater sentencing discretion for judges. It could thus be overridden by other considerations, such as if there was believed to be a significant danger of abuse of this discretionary power.

determinate conclusions regarding deserved sentencing. But they may provide some, albeit imperfect, guidance in the deliberations of those making sentencing decisions.

Second, even given that we do not know precisely what sentence is morally deserved in a particular case, the retributivist constraint on undeserved punishments can protect against focusing solely on consequentialist considerations in sentencing. As mentioned before, those who endorse retributivist constraints typically worry that, without such constraints, offenders might be given sentences well in excess of what they deserve to further some consequentialist aim. The retributivist requirement that punishments be only in a manner and to a degree that is deserved is not, in itself, able to ensure that political authorities never impose excessive punishments — at least, not without an account of moral desert that can ground specific sentencing determinations — but this retributivist constraint *is* sufficient to ensure that political authorities never impose excessive punishments *to further some consequentialist goal*. Note that this does not prohibit considerations of social benefits from playing *any* role in determining punishments; in Kantian terms, respect for humanity does not prohibit treating offenders as means to some consequentialist end, but rather it prohibits treating them *merely* as means to such an end.

The retributivist constraint, then, can tell us that sentencing determinations should be governed by considerations of moral desert, even if this constraint cannot itself ensure that the morally deserved sentence is established. This is a significant constraint. Again, the point here is not that it rules out certain sentences *per se* as inappropriate, but rather it rules out certain considerations as inappropriate in governing sentencing decisions. Of

course, this argument is unlikely to persuade a nihilist about moral desert. After all, if there is no such thing as a morally deserved punishment (or range of punishments) in particular cases, then it would make no sense to make moral desert one's governing aim. But if nihilism about moral desert is false (and I have suggested above that, at least, this remains an open question, and in fact our intuitions tend to oppose the nihilist), then the retributivist requirement that punishments be deserved can provide some useful guidance, at least in terms of what is the appropriate *aim* of sentencing (and, importantly, what aims are inappropriate).

So perhaps the retributivist consideration that punishments should not be more than is deserved can be of some use as a constraint on sentencing determinations. It may provide some guidance in deliberation about appropriate sentences in particular cases, and it can help to protect against excessive punishments inflicted solely on consequentialist grounds. But although this constraint can provide some useful guidance in determining how we may punish, it is insufficient. This is not, at least not primarily, for the reason suggested by critics such as Shafer-Landau, viz., that retributivism cannot make adequate sense of the notion of moral desert to ground determinate constraints on punishment. In fact, even if retributivism did provide the resources with which to determine definitively what punishment appropriately reflects the severity of (and offender's responsibility for) a crime, the retributivist constraint would still be insufficient. To begin to see why, I return to the *lex talionis*.

III. Retributivism and respecting offenders as moral persons

Shafer-Landau's critique of retributivism focuses on the difficulty of determining moral desert. But in discussing the strict, eye-for-an-eye version of *lex talionis*, one of his objections is worth noting. In addition to pointing out (as discussed earlier) that equal-in-kind punishment will be impossible in many cases, he alludes to a distinct objection:

Finally, in many cases where *lex* does offer concrete advice, many of the recommendations are morally unsavory — raping a rapist, or torturing a torturer, for instance. Some may be prepared to bite the bullet, and insist that such treatment is what these criminals deserve. Even if it is, we surely do not want such desert verdicts used as a basis for structuring sentencing guidelines. Better to knowingly fail to mete out such deserts than to authorize a line in the budget for an official rapist or torturer.³¹

Here the worry is not that *lex talionis* will be unable to provide definitive sentencing guidance; the worry is that it *will* be able to provide such guidance, and that we'll find its prescription morally unpalatable.

Recognizing this problem, Kant claims that certain exceptions to strict *lex talionis* will sometimes be warranted. For instance, for the crime of rape, Kant prescribes not that the perpetrator likewise be raped (he explains that this would itself be an unjustified violation of humanity), but rather that he be castrated. This punishment respects the law of retribution, Kant believes, "if not in terms of its letter at least in terms of its spirit."³² It is unclear why Kant believes castration is the appropriate, "like for like," response to rape. The physical damage of castration is not particularly similar to that of rape, nor is it

³¹ Shafer-Landau, "Retributivism and Desert," p. 193.

³² Kant, *Metaphysics of Morals* 6:363 (p. 498).

clear that the two are similar in their emotional impact (also, this punishment is inapplicable in cases of rape by a female). Instead, his discussion here feels *ad hoc*, as though he is trying to avoid the unsavory implication that a rapist deserves to be raped. But there seems to be nothing in the fairly straightforward, in-kind version of *lex talionis* that would warrant such a deviation.

It is just these kinds of unsavory implications of the strict *lex talionis* that have led many theorists to adopt alternative forms of retributivism, several of which I cited above. But all versions of retributivism face the general worry that, in many cases, the punishment determined on retributivist grounds to be morally deserved will be quite severe. Consider what the equal-suffering retributivist will be obliged to advocate as a deserved punishment for the torturer. Similarly, Nozick's equation for determining the appropriate retributive punishment (seriousness of crime multiplied by degree of responsibility) doesn't avoid the problem, at least not in cases where the rapist, or the torturer, acted fully voluntarily. In fact, any retributivist account, because it links moral desert to the seriousness of a crime, will face this problem. Some crimes are extremely heinous, and these are sometimes committed completely voluntarily. It's true that, in such cases, some may be inclined to endorse the extremely harsh punishment that retributivism will dictate. But to the extent that many of us are troubled by the idea of torturing the torturer, raping the rapist, etc., we will need to find some justification for restricting such practices.

One could, of course, simply insist that the punishment indicated by retributivism as the proper response to, say, torture is some significant degree of suffering that is

nevertheless not as severe as torture itself. Such an argument runs the risk, however, of appearing arbitrary or *ad hoc* unless the retributivist can offer some principled justification for the claim that the morally deserved response to the crime is less severe than the crime itself. It just isn't clear that retributivism has the resources to ground such an account. *Lex talionis* obviously cannot, nor can the equal-suffering account. The fair play account would need to explain why some less severe response is sufficient to remove the unfair advantage the offender gained by the more severe crime. And a communicative account such as Hampton's would need to explain why some less severe response can adequately nullify the message of the criminal's superiority to her victim(s). None of the standard retributivist accounts provides a clear explanation of why, as a response to some serious criminal wrongdoing, retributivism would indicate the infliction of some lesser degree of suffering on the perpetrator.

These concerns are, in part, why the retributivist constraint is typically cited in the form in its negative, or minimalist, form — that is, that punishment must be *only* as severe as is morally deserved. As indicated earlier, negative retributivism reflects the widespread intuition that offenders should not be punished more severely than their crimes deserve, but it also allows us to avoid the implication that we may be required to impose troublingly harsh punishments in certain cases. Thus it may seem that the appropriate remedy to the concerns about unpalatably harsh punishments raised by the *lex talionis* and other versions of retributivism discussed above is simply to endorse negative retributivist constraints.

Despite the intuitive appeal of negative retributivism, however, it's ultimately unclear how we could justify endorsing only this negative version, but not the positive version; at least it's unclear how we could justify this on purely retributivist grounds. Accounts of retributivism vary widely, but a general thread running through all accounts is that there is some sort of value in wrongdoers getting what they deserve³³ — either it is intrinsically valuable, or it is good because it communicates the appropriate message, or somehow nullifies the crime, etc.³⁴ Given that the value of wrongdoers' getting what they deserve is fundamental to retributivism, there seems to be no purely retributivist justification for distinguishing the view that offenders should get no *more* than they morally deserve from the view that offenders should get no *less* than they morally deserve, and for endorsing the former but not the latter. From a purely retributivist point of view, moral desert should determine both the ceiling and the floor for appropriate punishments. If there is no retributivist justification for accepting negative but not positive retributivism, however, then proponents of only the negative constraint will need to appeal to some nonretributivist considerations to justify endorsing one but not the

³³ As John Cottingham points out, "retribution" derives from the Latin *re+tribio*, "to pay back." Cottingham, "Varieties of Retribution," p. 238.

³⁴ Boonin, in *The Problem of Punishment*, contends that retributivism that is concerned with moral desert is only one form of retributivism. I disagree with Boonin's taxonomy, which, for instance, considers what I have called fair play theories not to be desert-based accounts. Developing a full objection to Boonin on this point is beyond the scope of this paper, but I will point out that the proponents of the fair play view take themselves to be addressing questions of moral desert — see, e.g., W. Sadurski, *Giving Desert its Due* (Dordrecht: Reidel, 1985); George Sher, *Desert* (Princeton: Princeton University Press, 1987); and Michael Davis, "Criminal Desert and Unfair Advantage: What's the Connection?" *Law and Philosophy* 12 (1993): 133-56. Regardless of whether there could be a retributivism that is not concerned with moral desert, however, my project in this paper is focused on the widely endorsed negative retributivist constraint that punishment should be no more severe than is morally deserved.

other. Retributivist considerations alone will thus be insufficient to ground constraints on punishment.

One might instead attempt to defend retributivist constraints by endorsing the full-throated retributivist claim that offenders should be punished to the full extent of their moral desert, but then insisting that this is only a presumptive principle, one that could be overridden by other concerns. This would allow for mitigation of punishments for various reasons, and thus provide a way to avoid the unwanted conclusion that extremely, unpalatably harsh punishments may in some cases be warranted. Whatever these overriding considerations happened to be, however, they would be nonretributivist ones. And so, again, on this strategy retributivist constraints would be insufficient.

Falls' retributivist account, discussed earlier, employs essentially this sort of strategy. On her Kantian theory, punishment is justified in that it holds wrongdoers morally accountable for their acts, and the suitable severity of punishment is that which appropriately communicates the state's message of moral condemnation. A punishment appropriately communicates this condemnatory message, on her view, when it is proportionate to the offense.³⁵ Thus her proportionality principle states that the "severity of punishment that is one's [earned moral desert] is the degree of severity proportionate to that of the wrongdoing."³⁶ Falls explicitly acknowledges that the morally deserved punishment will often be quite severe: "The proportionality principle is solely about

³⁵ Falls, "Retribution, Reciprocity, and Respect for Persons," p. 45.

³⁶ *Ibid.*, p. 41.

earned moral desert, and according to it torture, death, whatever, can be the earned moral desert of the most wicked.”³⁷

Falls’ proportionality principle turns out to be only presumptively binding, however. In cases in which one’s earned moral desert would be torture or death, etc., the proportionality principle will be overridden by what she calls a limiting principle. The limiting principle states that “punishment is justified only if the one suffering it remains capable of reflectively responding to the treatment being received and the condemnation it communicates.”³⁸ Whereas the proportionality principle is grounded in considerations of *earned* moral desert, the limiting principle is grounded in considerations of *unearned* moral desert: As Falls sees it, an implication of Kant’s principle of respect for humanity is that “persons simply as persons deserve that the state hold them morally accountable (at least for certain kinds of acts).”³⁹ This desert is unearned — that is, each of us deserves to be held accountable simply in virtue of our humanity. But to hold a person accountable, or responsible, for wrongdoing requires that we allow the offender to respond to her punishment (to the message it communicates) as a moral agent. Thus punishments that preclude such response are prohibited.

I believe Falls is onto something here. But note that her limiting principle, as attractive as it is, is not a retributivist constraint. Falls makes clear that retributivist claims

³⁷ Ibid., p. 47.

³⁸ Ibid. The proportionality principle is presumptively binding because even when it is overridden by the limiting principle, it remains the case that offenders still morally deserve (in the sense of earned moral desert) the particularly harsh response.

³⁹ Ibid., p. 41.

such as her proportionality principle are earned-moral-desert claims.⁴⁰ Her limiting principle, by contrast, is justified by considerations of unearned moral desert. It is clear, then, that the limiting principle is not a retributivist principle. So again, on this account retributivism is insufficient to guide how a political authority may punish; an additional, nonretributivist constraint is warranted.

I actually find Falls' limiting principle quite plausible. It helps to clarify that the treatment an offender morally deserves may be a matter of more than the severity of her crime and her degree of responsibility for it. Retributivism, grounded in the notion of payback, is insufficient in its inability to account for the fact that the treatment an offender deserves is a matter not only of the crime she has committed, but of who she is. As Falls puts it, part of what each of us deserves is unearned — that is, it attaches merely in virtue of our status as moral persons.⁴¹ Retributivism, however, deals only in earned desert, i.e., the treatment an offender comes to deserve through the commission of her crime. Thus retributivism is unable to make sense of the intuition many will share that an offender does not always deserve to be treated as harshly as might be indicated by considerations solely of the seriousness of the crime and her degree of responsibility for it.

⁴⁰ Ibid., p. 40.

⁴¹ Stephen Darwall makes a similar point in "Two Kinds of Respect," *Ethics* 88:1 (1977): 36-49.

IV. Conclusion

Retributivism is unable to account for the idea that the treatment an offender deserves all things considered is not solely a matter of the treatment she has earned by committing a crime. Part of what an offender deserves is grounded in her status as a moral person. In the next chapter, I defend an additional constraint on punishment, one grounded in the Kantian notion that a central feature of humanity is our capacity for moral reform and redemption. Kant believed that redemption was always possible, even for the most vicious individuals. Thus he wrote:

[T]he censure of vice ... must never break out into complete contempt and denial of any moral worth to a vicious human being; for on this supposition he could never be improved, and this not consistent with the idea of a *human being*, who as such (as a moral being) can never lose entirely his predisposition to the good.⁴²

Building on Kant's discussions of contempt and the prospect of redemption, I contend that punishments should not, in their mode or degree, tend to undermine offenders' prospects of moral reform, and that certain forms of punishment, notably capital punishment and incarceration in certain types of maximum-security facilities, tend to do just this. In doing so, they fail to respect offenders as persons, as ends in themselves.

My goal in this chapter, however, has been to lay the groundwork for this upcoming account by demonstrating the insufficiency of the commonly cited retributivist constraint. Contrary to those who object that retributivism is either useless or false

⁴² Kant, *Metaphysics of Morals* 6:463-64 (p. 580)

because it is unable to provide definitive guidance regarding what punishments are morally deserved, the retributivist constraint can provide some genuinely useful guidance. Nevertheless, the guidance retributivism does offer will in some cases strike many of us as excessively, inflexibly harsh. Retributivism itself provides no reason to accept its negative prescription (punishments no more severe than is deserved) while rejecting its positive one (punishments no less severe than is deserved). And by focusing only on what an offender morally deserves *for her crime*, (which may in some cases be quite severe), retributivism fails to account for what an offender morally deserves all things considered.⁴³

⁴³ I am grateful to Larry May, Christopher Heath Wellman, and David Wood for their helpful comments on previous drafts of this chapter. Also, I presented a version of this paper at the Washington University in St. Louis Workshop on Politics, Ethics, and Society in November 2008. I appreciate the thoughtful feedback I received from participants at that workshop.

CHAPTER 4

Punishment, contempt, and the prospect of moral reform

For the institution of criminal punishment to be morally permissible, it is widely accepted that penal practices must treat criminals with some basic level of respect. The most common account — endorsed (whether explicitly or implicitly) by most theorists of punishment and widely reflected in legal practice — is that respect for offenders requires that punishments not be more severe than the corresponding crimes morally deserve. In chapter 3, I contended that this retributivist constraint, although a valuable check against the imposition of excessive punishments on consequentialist grounds, is ultimately insufficient for two reasons: First, as has been widely noted, it is often unclear precisely what punishment (or range of punishments) is morally deserved in a given case. Second, and more importantly, what guidance we do get from prominent retributivist accounts indicates that the punishment that is morally deserved for a given crime may often be quite severe — so severe, in fact, as to strike many as intuitively excessive. This is in part because retributivism seeks to assess desert solely in terms of what the offender did, rather than who she is, or who she can be. The concerns associated with retributivism give us reason to examine more closely what treating offenders with respect requires, and whether respect warrants some additional constraint(s) on punishment.

My aim in this chapter is to argue for one such constraint, albeit one that has often been overlooked both in the literature on punishment and in legal practice. Specifically, I contend that punishment, if it is to treat offenders with respect as autonomous moral

persons, should never be of a mode or to a degree that tends to undermine the prospect of offenders' moral reform. As I discuss below, although this principle is centrally concerned with the moral reform of the offender, my account differs from typical offender-improvement theories of punishment, insofar as such accounts characteristically set offender improvement as the aim of punishment, whereas on my account reform need not be an aim of punishment but is rather a prospect that penal practices should take care not to undermine.

In section I, I appeal to certain discussions in Kant, not (primarily) of respect but rather of contempt, which Kant took to be the opposite of respect. By highlighting certain relevant features of contempt that make it especially problematic, we can gain insight into what is, conversely, required by respect for persons. After discussing what is implied by contemptuous treatment generally, I consider in section II what would constitute contempt, and thus would violate the requirement of respect, in the context of punishing criminal offenders. Again, on my account, respect prohibits punishments that tend to undermine offenders' prospects for moral reform. In sections III and IV, I further flesh out my own view, first by contrasting it with accounts that cite reform as a positive aim of punishment, and then by considering some possible objections to my account.¹

¹ I should note at the outset that, although my account takes as its starting point certain Kantian themes, my central project here is not one of Kant interpretation. My aim is to consider what conclusions we might draw if we apply certain useful, compelling themes in Kant's practical philosophy to an analysis of how a political community may permissibly treat offenders. In fact, certain conclusions that I reach about which punishments are and are not permissible (e.g., regarding capital punishment) contradict Kant's own explicit conclusions. But whether or not Kant's explicit about punishment can be reconciled with certain other, fundamental principles in his moral philosophy, my aim is to develop and expand on certain Kantian themes that I take to be particularly plausible and to suggest how they are relevant to questions of

I. What's wrong with contempt?

Kant famously believed that all human beings, as rational beings, possess a dignity, an absolute inner worth, that warrants respect (both self-respect and the respect of others).² Thus his second formulation of the categorical imperative instructs us, “Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end, never simply as a means.”³ In various passages Kant indicates that the converse of respect, as he conceives of it, is contempt. In the *Doctrine of Virtue*, for instance, he writes, “To be contemptuous of others (*contemnere*), that is, to deny them the respect owed to human beings in general, is in every case contrary to duty; for they are human beings.”⁴ One strategy, then, for better understanding what respect for persons requires is to examine, conversely, what it prohibits, by looking more closely at what it means to treat others with contempt.

punishment. Thus, my concern is less with whether Kant's theory does, or should, commit him to the conclusions I draw than with whether mine is a plausible account in its own right.

² See, e.g., Kant, *Metaphysics of Morals* 6:435 and 6:462. In the Cambridge Edition of the Works of Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary J. Gregor (Cambridge, U.K.: Cambridge University Press, 1996), pp. 557 and 579, respectively. All citations to the *Metaphysics of Morals* are to this edition.

³ Kant, *Groundwork of the Metaphysics of Morals* 4:429. In Immanuel Kant, *Ethical Philosophy*, 2d ed., trans. James W. Ellington (Indianapolis, Ind.: Hackett Publishing, Inc., 1994), p. 36. All citations to the *Groundwork* are to this edition.

⁴ Kant, *Metaphysics of Morals* 6:463 (p. 579). Similarly, in Georg Ludwig Collins' notes from Kant's ethics lectures, Kant explicitly describes contempt as the opposite of respect. In the Cambridge Edition of the Works of Immanuel Kant, *Lectures on Ethics*, trans. Peter Heath and ed. Peter Heath and J.B. Schneewind (Cambridge, U.K.: Cambridge University Press, 1997), p. 173.

To treat someone with contempt, on Kant's view, is to treat her as morally worthless.⁵ This conception may strike some as too strong. Michelle Mason, for instance, describes contempt somewhat more modestly, "as presenting its object as low in the sense of ranking low in worth as a person in virtue of falling short of some legitimate interpersonal ideal of the person."⁶ Thus for Mason, contempt takes its object to be fundamentally deficient, or comparatively low in worth, along some dimension(s) of moral personhood, but perhaps not altogether morally worthless. Whether contempt picks out a (perceived) complete lack of moral worth in someone, however, or just a fundamental deficiency, we can say that a person regards the object of her contempt as inferior as a person.

It's worth noting here the distinction between *treating* someone with contempt and *regarding* her with contempt. Kant recognized this distinction, and he seemed to view contemptuous regard as, at least to a degree, outside our control. He writes, "At times one cannot, it is true, help inwardly looking down on some in comparison with others; but the outward manifestation of this is, nevertheless, an offense."⁷ Here Kant indicates that contempt as regard is, or significantly involves, a feeling that cannot be rationally controlled at the time at which it emerges.

Still, it's plausible that one could, over time, cultivate a disposition not to regard others with contempt. To the extent that this is possible, there are at least two good

⁵ See, e.g., *Metaphysics of Morals* 6:462 and 6:463-4 (pp. 579-80).

⁶ Michelle Mason, "Contempt as a Moral Attitude," *Ethics* 113 (January 2003): 234-272, pp. 240-41.

⁷ *Metaphysics of Morals* 6:463 (pp. 579-80).

reasons to do so: First, it may be exceedingly difficult to harbor contemptuous feelings for another without these feelings manifesting themselves in contemptuous treatment. Thus if contemptuous treatment is morally prohibited, then as a practical matter, meeting this moral proscription may require cultivation of a noncontemptuous disposition. Second, even if one could conceal one's contemptuous regard, such regard veiled with apparently respectful treatment seems disingenuous, and thus it may not actually be respectful at all. Imagine discovering that a colleague who has always treated you respectfully has actually at the same time regarded you with contempt. On finding this out, it might cross your mind, in retrospect, that the years of seemingly respectful treatment actually reflected the depth of the colleague's contempt: She didn't even regard you as worthy of her honesty with respect to her assessment of you as a person. Contemptuous regard may thus be troubling even when it does not manifest itself in obviously contemptuous treatment. For both of these reasons, then, even if contemptuous regard does not admit of rational control at the time it surfaces, there are reasons to work to cultivate a disposition against contemptuous regard. And although the focus in what follows is on contemptuous treatment, contemptuous regard will at times be relevant to the discussion as well.

A relevant feature of contempt, evident in the previous description of it as regarding its object as inferior as a person, is that it is person-focused rather than act-focused. Mason rightly distinguishes contempt from resentment in that, whereas resentment is typically focused on what someone has done or brought about ("I resent that she unfairly embarrassed me in front of our colleagues"), contempt focuses on the person — not for what she has done, but for who she is. Thus Mason writes:

Contempt, to adopt a phrase of Augustine's, thus will have none of "Despise the sin but not the sinner." The "sin" in such a case is simply an outer manifestation of something taken to go to the core of the "sinner," something taken to be contemptible.⁸

Furthermore, not only is contempt person-focused, but it is pervasively person-focused.

In other words, contempt permeates all of our interactions with those we hold in contempt. As Mason puts it, my contempt for another becomes that person's "most salient description for purposes of my ... assessment of her."⁹ So whereas my resentment for something a person has done may not necessarily color my entire evaluation of (and all my interactions with) the person, contempt tends to present a person as morally inferior generally, or at least in her most fundamental aspects (viz., in the ways we take to matter most).

Just as distinguishing contempt from resentment highlights certain relevant features of contempt (namely, its person-focus and its pervasiveness), it is also instructive to consider how contempt contrasts with anger. Whereas anger is heated, contempt is cold. An angry response engages the other for a perceived offense; a contemptuous response, on the other hand, treats the other as not worth the trouble. Thus Kant states:

[W]e cannot be angry and at the same time hold the other in contempt; for just as anger involves an emotion that presupposes a great exertion of effort to resist the impression of a felt offence, so contempt incorporates a conviction of the object's unworthiness for employment of

⁸ Mason, "Contempt as a Moral Attitude," 247.

⁹ Ibid., 249.

such a resistance on its behalf, and is therefore coupled with calmness.¹⁰

Contempt is distinctive, then, in that whereas other negative responses to someone — such as anger, censure, or heated argument — treat their object, at least implicitly, as worthy of concern and engagement as a moral person, contempt does not. Of course, we might still treat such a person as worthy of our prudential concern, perhaps as a threat. But assessing someone as worthy of prudential concern is consistent with maintaining a moral attitude of contempt toward her, with regarding and treating her as beneath our moral concern. To treat someone as *solely* of prudential concern would be to treat her merely as a thing, rather than as a moral person — in Kantian terms, it would be to treat her merely as a means rather than as an end.

Contempt, then, presents a person as fundamentally deficient (if not altogether worthless), is focused on the person rather than what she has done or brought about, pervades our entire assessment and interaction with her, and is cool and dismissive in virtue of not regarding her as worthy of engagement as a moral person.¹¹ Another feature of contempt, which follows from these, is that it is especially uncondusive to the prospects of moral reform, forgiveness, and reconciliation; in fact, it is arguably less

¹⁰ From Johann Vigilantius' notes from Kant's ethics lectures. In the Cambridge Edition of the Works of Immanuel Kant, *Lectures on Ethics*, trans. Peter Heath and ed. Peter Heath and J.B. Schneewind (Cambridge, U.K.: Cambridge University Press, 1997), p. 417. See also p. 434 and p. 173 (the latter is from Collins' lecture notes).

¹¹ Note that, although in this chapter I write of contempt as directed at others, the points I mention here also apply to self-contempt.

compatible with these ideals than are the hotter, more engaged emotions such as anger or even hatred.¹² As Thomas Hill writes:

[C]ontempt is a deep dismissal, a denial of the prospect of reconciliation, a signal that conversation is over. Furious argument and accusation, and even sharp-tongued deflation of hypocrisy and self-deception, leave some space to resume communication; but cold, silent contempt does not. The one demands to be heard, while the other walks away in disgust.¹³

As Hill indicates, we can see reflected in contempt a sort of disengagement, a giving up on a person. Or similarly, contempt may reflect one's never regarding the person as worthy of genuine engagement in the first place.

Because contempt essentially reflects a giving up on its object, it will typically be unresponsive to evidence of moral reform — that is, evidence that might count in favor of forgiveness. After all, the relative calmness characteristic of contempt reflects not only a belief that the person is not worth our engagement, but also a sort of confidence that the person's status as fundamentally subpar is settled, and not liable to change.¹⁴ And

¹² Perhaps not surprisingly, psychologist John Gottman has found that contempt between spouses is the most significant predictor of divorce. See John M. Gottman, James D. Murray, Catherine C. Swanson, Rebecca Tyson, and Kristin R. Swanson, *The Dynamics of Marriage: Dynamic Linear Models* (Cambridge, Mass.: MIT Press, 2002). For a discussion of Gottman et al.'s findings, see Malcolm Gladwell, *Blink: The Power of Thinking Without Thinking* (New York City: Little, Brown and Company, 2005), pp. 18-33.

¹³ Thomas E. Hill, *Respect, Pluralism, and Justice: Kantian Perspectives* (Oxford, U.K.: Oxford University Press, 2000), p. 60.

¹⁴ One might think that insofar as contempt is cool and calm relative to anger, hatred, etc., it would thus be more conducive to our recognizing evidence of reform. But as I indicate above, the calmness characteristic of contempt is a product of our being settled in our assessment of the other person's moral character. Thus although contempt is calm, it isn't necessarily conducive to the sort of calm, cool deliberation that (for instance) Hume endorsed, because the calmness of contempt comes essentially from one's having finished deliberating, from having written the other person off, so to speak.

because contempt is characteristically unresponsive to evidence that might favor forgiveness, it will also typically be incompatible with the prospect of reconciliation.

Mason, who in her essay aims to defend the place of contempt as a morally justified attitude, claims that contempt can remain responsive to evidence that would count in favor of forgiveness. Her argument, however, is ultimately unconvincing. Contempt, on Mason's conception, assesses its object as low according to some interpersonal ideal. But as Mark Kalderon writes, "Even if someone were lacking in this way, to treat him as an end is to treat him as capable, at least in principle, of acquiring the requisite sensitivity and perception. ... The difficulty of course is that contemptuousness is inconsistent" with such treatment.¹⁵ In response to this sort of worry, Mason contends that it is both empirically and conceptually plausible that contempt may be sensitive to evidence in favor of forgiveness.¹⁶ She writes:

In response to the empirical claim, I have only my own experience as counterexample and ask others to consult experiences of their own. As for the conceptual claim, I do not see the basis for it; common usage, for example, does not suggest that it is part of the very meaning of contempt that once one is a contemner, one is forever a contemner.¹⁷

In my view, Mason's reply here is unpersuasive. Undoubtedly, many of us can recall instances of having held someone in contempt only to have that assessment change when evidence favoring forgiveness emerged. But presumably there also are likely to have been many counterexamples, in which our contempt rendered us slow to recognize

¹⁵ Mark Kalderon, *Moral Fictionalism* (Oxford, U.K.: Oxford University Press, 2005), 176.

¹⁶ Mason, "Contempt as a Moral Attitude," 256.

¹⁷ *Ibid.*

mitigating evidence, or even prevented our recognizing it altogether. Thus even if we can recall cases in which contempt did not so color our view of a person as to leave us unable to recognize evidence of repentance or reform, this does not dispel the empirical worry that contempt *tends* to have this practical effect.¹⁸

Furthermore, the claim that contempt renders us less sensitive to evidence favoring forgiveness is supported by certain other features of contempt discussed above. I followed Mason in describing contempt as person-focused and pervasive. That is, I regard someone with contempt not because of particular actions by which she may have wronged me, but rather for who she is as a person. Also, my contempt colors my entire assessment of and interaction with her.¹⁹ Given this conception, what sort of changes might I regard as evidence in favor of forgiveness? Presumably the most viable recourse for the person held in contempt, if she is genuinely repentant and desires forgiveness, is to apologize and change her behavior. But given that my contempt transcends any particular harms or transgressions toward me, why expect that particular acts of kindness or contrition would elicit the withdrawal of that contempt? Since my contempt permeates my assessment of the person, I would more likely view any sincere acts of contrition as disingenuous, or perhaps as somehow manipulative. Again, this is not to say that

¹⁸ Note that it does not help to stipulate that contempt is morally justified only in the cases in which it does not leave one unable to recognize mitigating evidence. For contempt is contempt in either case; insofar as it *tends* to render people unable to see mitigating evidence, it is implausible to condemn contempt only in those cases in which this ever-present tendency happens to be realized.

¹⁹ Perhaps a useful, albeit imperfect, analogy could be drawn here between contempt and love. Love, we often say, can tend to blind us to (or at least leave us less sensitive to) the flaws in those whom we love. My argument is that contempt tends to have a similar affect, although in the opposite direction: Contempt tends to blind us (or at least leave us less sensitive to) the redeeming qualities in those whom we regard with contempt.

contempt *always* prevents us from recognizing evidence of reform, or that the concept of contempt itself rules out this possibility. But certain central features of contempt make it unlikely that contempt will typically leave us capable of recognizing evidence in favor of forgiveness. Contempt, after all, does not look for signs of repentance; contempt gives up the search.

There is another sense in which contempt is in tension with reform. Not only does contemptuous regard tend to leave us unresponsive to evidence of reform, but also contemptuous treatment may tend to undermine the prospect of reform. This may occur in at least two, related ways: First, an implication of our treating a person as though she is fundamentally morally inferior, of dismissing her as not being worth our continued engagement, is that we will not be motivated to provide her with the resources and opportunities that might facilitate her reform. I say more about this below, in the context of whether penal institutions make available sufficient opportunities for offenders, should they so choose, to help themselves. Also, when we treat a person as though she is fundamentally inferior, when our treatment of her expresses that we have given up on her, that she is not worth our continued engagement and effort, she may in time come to agree with this assessment and, as a consequence, give up on herself. Thus again, our contempt may not only blind us to evidence of reform, but it may also actually undermine the prospect of reform.

Given this conception of contempt, we can perhaps see why Kant regarded contempt as “in every case contrary to duty.”²⁰ In particular, Kant was concerned about the tension between contempt and reform. He writes,

[the] censure of vice ... must never break out into complete contempt and denial of any moral worth to a vicious human being; for on this supposition he could never be improved, and this [is] not consistent with the idea of a *human being*, who as such (as a moral being) can never lose entirely his predisposition to the good.²¹

As this passage indicates, Kant saw contempt as inconsistent with the recognition of humanity’s predisposition to morality, and thus with a recognition of the possibility of redemption.

Much more could be written about contempt, but for my purposes the salient features are these: Contempt is person-focused and pervasive; it presents its object as inferior, if not altogether worthless, as a moral person; and it is cold and dismissive, i.e., it essentially reflects our giving up on the person as a moral person. In virtue of these features, contempt tends to be unresponsive to evidence of repentance and reform, and it actually tends to undermine the prospect of reform as well.²²

²⁰ *Metaphysics of Morals* 6:463 (p. 579).

²¹ *Ibid.* 6:463-4 (p. 580).

²² I acknowledge that some may use the term “contempt” in ways that diverge from the conception I have developed here. For instance, some may on occasion use “contempt” as roughly analogous to “hatred,” as burning hot, rather than as cold. Or some may use “contempt” as more closely analogous to “resentment” in the sense of being act-focused. Although I believe the conception of contempt that I have developed is intuitively plausible, it is undoubtedly the case that various individuals use the term “contempt” in somewhat varying ways on various occasions. (William Ian Miller, for instance, suggests a much more inclusive, and innocuous, conception of contempt, one that has “a light side as well as a dark one,” according to which we may regard our children, pets or even on occasion our partners with contempt insofar as we judge them to be endearingly subordinate and unthreatening. Miller, *The Anatomy of Disgust* (Cambridge, Mass.: Harvard University Press, 1997), p. 32.) For my purposes, however, it is enough if

By contrast, then, what does respect for persons require? It may require quite a bit, and my aim here is not to give a full account of what respect demands.²³ Given the features of contempt that I have discussed above, however, certain requirements of respect become apparent. Respect requires not adopting a dismissive attitude toward others, not giving up on them as irredeemably deficient as moral persons. It follows, then, that respect requires us to be open to evidence of repentance and reform, and never to treat others in ways that will tend to undermine the prospect for their reform. In the next section, I consider what implications these conclusions will have for how the practice of criminal punishment should be constrained.

II. Treating offenders with respect

Given the conception of contempt that I have endorsed above, and thus, conversely, what I have argued is required by respect for persons, my central argument with regard to punishment follows fairly straightforwardly: On my view, for punishment properly to respect offenders as persons, it should not express contempt for them; it shouldn't give up on them, so to speak. In this section, I argue that this prohibition on contempt in our treatment of offenders grounds the following constraint on punishment:

there is form of regard and treatment that exemplifies the features I have discussed in this section. I claim that this is most naturally described as contempt. But what it is called is less important than that we can recognize a form of regard and treatment that exhibits the features I have suggested — and, as I argue below, that our existing penal practices often do (but should not) express this attitude in their treatment of offenders.

²³ For an excellent discussion of respect in the context of moral censure, see Hill, *Respect, Pluralism, and Justice: Kantian Perspectives*, pp. 114-18.

Punishments should not, in their mode or degree, tend to undermine the prospect of offenders' reform.

Essentially, punishment undermines the prospect of reform when it imposes conditions on offenders such that the punishment itself makes it less likely (than had the punishment not been imposed) that they will engage in moral reflection about the wrongdoing for which they are being punished, come to see their actions as morally wrong, regret having done them and, consequently, make a genuine commitment to change their behavior in the future. The imposition of such conditions, on my account, reflects a lack of respect for offenders as persons. Punishments may impose such conditions in various ways.

First, punishment may undermine the prospect of reform when a sentence is so severe that it inhibits the criminal's capacity to engage in the moral reflection necessary to come to see her criminal behavior as wrong.²⁴ Capital punishment is an obvious example of such a sentence, in that execution clearly inhibits (in fact, extinguishes) the individual's capacity for moral reflection. One might object here that a criminal sentenced to death may have substantial time to engage in moral reflection, should he so choose, while he waits on death row during his appeals process. The appeals process, however, should not be conflated with the sentence; the death penalty itself, when carried out, undeniably extinguishes a person's capacity for moral reflection.²⁵ In doing so, it

²⁴ C.f., M. Margaret Falls, "Retribution, Reciprocity, and Respect for Persons," *Law and Philosophy* 6:1 (April 1987): 25-51, esp.pp. 46-48.

²⁵ This is true unless, of course, one believes in an afterlife in which moral reflection is still possible. If so, one will presumably deny that execution extinguishes this capacity. For a number of reasons, however,

fails to afford the offender the respect appropriate to him as a person. Similarly, punishments imposing extreme physical suffering would be prohibited insofar as the severe pain associated with such treatment would significantly inhibit the individual's capacity for moral reflection.

Punishment may inhibit the capacity for moral reflection not only because of the severity of the sentence itself, but also because of the conditions that prevail in the carrying out of that sentence. Two criminals receiving identical prison sentences may nevertheless be said to receive different punishment if one is subject to significantly harsher conditions during his incarceration than the other (e.g., physical beatings, rape, etc.). Much of the philosophical literature on punishment, to the extent that it addresses the question of how (as opposed to why) we should punish offenders, centers on morally justifiable sentencing practices. Focusing only on imposing justified sentences, however, without also considering the conditions in which the sentences are administered, overlooks how much of the work of determining the overall character of a particular instance of punishment is done after the sentence is handed down.²⁶ Respect for offenders must therefore be expressed not only in the formal sentences themselves, but also in the manner in which these sentences are carried out.²⁷ Thus, the constraint I endorse here

religious claims of this sort are not promising bases for the establishment of penal practices in a liberal polity.

²⁶ C.f., "Gently Does It; Prison Conditions," *The Economist*, U.S. edition, July 28, 2007.

²⁷ Kant alludes to this point in the context of capital punishment, when, after endorsing capital punishment as the required sentence for murder, he nevertheless insists that the execution "must still be freed from any mistreatment that could make the humanity in the person suffering it into something abominable." *Metaphysics of Morals* 6:333 (p. 474). Unlike Kant's explicit claim in this passage, of course, my account entails that capital punishment is not morally permissible.

prohibits not only sentences that tend to undermine the prospect of offenders' reform, but also the toleration by penal institutions of conditions in the carrying out of sentences that tend to undermine the prospect of reform.

Another way punishments may undermine the prospect of reform is by weakening offenders' motivation, rather than their capacity, to undertake a process of moral reflection and reform. (As before, the weakening can be a feature either of the sentence itself or of the conditions involved in the administration of the sentence.) First, when a sentence expresses to an offender that society has given up on him, that it regards him as irredeemable, he may come to accept society's judgment about him, to see himself the same way. Thus his motivation to engage in the process of moral reflection and reform may be weakened. Second, if the punishment is so harsh, demeaning, etc., as to foster hatred (or perhaps cold, dismissive contempt²⁸) in the offender for the penal institution and the political community on whose behalf it punishes, then punishment may in this way weaken an offender's motivation to contemplate whether his criminal actions were, in fact, wrong. The laws of a political community, after all, can be seen as reflecting, and even expressing, that community's values.²⁹ Thus if an offender, because of the harsh or humiliating treatment he receives from the community's penal institutions, comes to regard the community itself with hatred or contempt, then he may likewise be less

²⁸ The notion of contempt directed upwardly — that is, from those out of power toward those in power — is discussed at length by William Ian Miller in "Mutual Contempt and Democracy," in his *The Anatomy of Disgust*, supra. n. 21, at pp. 206-34.

²⁹ For a useful discussion of the law as an expression of the values of a political community, see R. A. Duff, "Liberal Legal Community," in his *Punishment, Communication, and Community* (New York City: Oxford University Press, 2001), pp. 35-73. See also Igor Primoratz, "Punishment as Language," *Philosophy* 64 (1989): 187-205, esp. pp. 196-98.

motivated genuinely to reflect on which of the community's values were expressed in the law(s) that he violated, and whether these values might in fact have merit. Such punishment would therefore make reform less likely for the offender.

In addition, when a punishment does not at least make available the opportunities by which an offender may improve his situation, both during his punishment and afterward, this may weaken his motivation to engage in a process of reflection that may lead to repentance and reform. Here it is useful, following R. A. Duff, to distinguish "reform" from "rehabilitation," where "reform" refers to a change in individuals' motives and dispositions and "rehabilitation" refers to the improvement of their skills, capacities and opportunities.³⁰ Thus, on this construal, providing an offender with the education, training or information he needs to improve his chances of (re)integrating himself as a productive member of society would constitute an example of rehabilitation. Reform, however, would require something different, namely, that the offender come to appreciate that what he did was wrong and make a commitment to change his behavior accordingly.

Although the two concepts are distinct, they are nevertheless related. When penal institutions fail to make available to offenders the resources that could facilitate their reintegration into society, they send a message that society sees these offenders either as incapable of improvement, as not worth the trouble, or both. Also, apart from this expressive function (that is, even if the offender doesn't recognize the contemptuous message being implicitly expressed), failures to make such resources available may still

³⁰ Duff, *Punishment, Communication, and Community*, p. 5.

facilitate self-contempt in offenders. Offenders with few or no job skills, little or no education, and, importantly, no real tools with which they stand a serious chance of addressing these deficiencies and improving their lives may come to regard themselves as hopelessly incapable of improving their situations. That is, they may be more likely to regard themselves with contempt.³¹ Thus an offender's motivation to reform may be weakened by the message sent by a punishment's failure to make available sufficient opportunities for improvement, or by the lack of opportunities itself, or both. In any case, the result will be effectively to undermine the prospect of reform, and thus to fail to respect the offender as a person.

As an example of a form of punishment that exhibits a number of the tendencies discussed here, consider control units, or maximum security units. These are facilities in which prisoners typically spend twenty-three or more hours each day in isolation in their cells, exiting only to shower or for solitary exercise in a small yard. In one sense, such units might seem especially conducive to the sort of moral reflection that might lead to reform. The extreme isolation of such units, after all, offers prisoners plenty of time (in fact, little else) during which they may choose to contemplate the wrongness of their acts. In practice, however, evidence indicates that prolonged terms of isolation may undermine

³¹ Because the contrast class on this account is the likelihood of reform *had the offender not been punished*, one might raise this objection: Many offenders face considerable hardships, in particular a lack of opportunities to improve their condition, in their normal lives, outside of punishment. Given such conditions, the practices I have targeted here (i.e., lack of available resources, and a message of society having given up on the individual) might not really make reform less likely, given the likelihood that these resources would be lacking, and the expression of societal contempt would be present, in the offender's life even if she were not punished. As I see it, however, this is a problem not for my constraint on punishment, but for societal conditions generally. Perhaps my proposal could better be read to say that punishments should not undermine the prospect of reform in the sense of making reform less likely than it *should have been* had the offender not been punished.

both the capacity and motivation for moral reflection, and may instead push many prisoners toward paranoia, hopelessness, or desperation. Lorna Rhodes, an anthropologist who visited and studied control units in the state of Washington, writes that in “isolation or semi-isolation, there is nothing to nudge the mind outside of its self-preoccupation and discomfort.”³² Among the numerous prisoners she quotes, one says: “There is no hope for my future, no matter how hard I try to just be patient, be humble ... There is nobody to talk to ... and vent my frustration and as a result, sometimes I am violent. Pound on the walls. Yell and scream.”³³

In addition, Rhodes describes how prison officials are trained to regard prisoners skeptically, to view positive responses by prisoners not as good behavior but as waiting and manipulation. As one administrator put it:

The inmates know the game. They know what to say ... you have to be really guarded on that. You'd like to think that they're not animals, that these are human beings: 'Oh, my gosh, [he] won't do that. And he even promised me.' But that is not the reality ... there's always the risk that the person is just playing the game.³⁴

Undoubtedly, many of the individuals detained in maximum security units are adept at “playing the game.” But notice how the administrator’s description reflects the feature of contempt I discussed earlier, namely, its tendency to be unresponsive to evidence of moral reform. Officers in the control units are trained to regard prisoners as incapable of

³² Lorna Rhodes, *Total Confinement: Madness and Reason in the Maximum Security Prison* (Berkeley, Calif.: University of California Press, 2004), p. 111.

³³ *Ibid.*, p. 112.

³⁴ *Ibid.*, p. 167.

changing, and to discount apparent evidence of change as in fact evidence only of manipulation. As an officer told Rhodes: “A person is not a liar because he lies, but he lies because he’s a liar. The point is, how do we remove the liar out of a person? We can postpone lying ... but that does not change the individual.”³⁵ Certainly, many prisoners can be dangerous and manipulative, and the officers who deal with them on a regular basis should be wary. But to presume that each prisoner’s moral character is settled, and that any apparent signs of improvement must instead be evidence of manipulation, is essentially to give up on these prisoners. In my view, this essentially amounts to a form of institutionalized contempt.

Ultimately, Rhodes is pessimistic about the effects of prolonged isolation on prisoners’ moral (and mental) well-being. This is not to deny that in some cases isolation might facilitate moral reflection and reform. But the effects of prolonged isolation merit further study. If empirical evidence indicated that such punishments, on balance, did tend to hamper offenders’ ability to engage in moral reflection, and thus tended to undermine the prospect of their moral reform, then these punishments would be prohibited according to the constraint I am endorsing.

III. Reform to be promoted or not undermined?

On the conception of contempt I have endorsed, a central characteristic is that it expresses a giving up on its object. Respect, conversely, requires that we not treat others

³⁵ Ibid., p. 171.

as hopelessly, irredeemably inferior as moral persons; that is, respect requires not giving up on others. One might expect, given this account, that I would defend a different constraint on punishment, namely, that to be permissible, punishments must tend to *promote* the moral reform of offenders. One might think, after all, that the best way to express that we have not given up on an individual is to make her reform our positive aim, or goal. Thus perhaps our institution of punishment should reflect this positive goal, tailoring penal practices so that they tend to promote the moral reform of offenders (rather than merely the negative goal, on my account, of not tending to make offenders' reform less likely).³⁶

Typically, accounts of punishment that focus on considerations of reform of the offender do incorporate these as positive goals to be promoted. Prior to the 1970s, for instance, offender improvement was a prominent justification of punishment. R. A. Duff and David Garland write:

The same programmes and attitudes which fostered 'the Welfare State' sought to make the penal system an instrument of social engineering through which crime could be prevented. Punishment could prevent crime by deterring potential offenders or by incapacitating actual offenders: but it could achieve even greater goods, it was hoped, by reforming and rehabilitating offenders.³⁷

³⁶ The goal of offender improvement is reflected in the typical description of prisons as "correctional facilities." In the United States, all but two states incorporate "correction" into the name of the departments charged with managing their prison systems. The two exceptions are Hawaii (Department of Public Safety) and Texas (Department of Criminal Justice). It's unclear, however, how many of these state institutions actually regard offender improvement as central to their mission.

³⁷ R. A. Duff and David Garland, "Introduction: Thinking about Punishment," in Duff and Garland, eds., *A Reader on Punishment* (Oxford, U.K.: Oxford University Press, 1994), pp. 8-9.

As Duff and Garland explain, enthusiasm for offender improvement as a goal of punishment decreased significantly beginning in the 1970s, for two general reasons. Both lines of objection are compelling against offender improvement as a positive goal, but neither is damaging to my account of reform as a prospect not to be undermined. The first general objection centers on efficacy. Emerging evidence seemed to indicate that the penal programs of the mid-20th century aimed at reform simply didn't work. Insofar as an account grounds punishment's permissibility in its potential as a mechanism for reforming offenders, evidence casting doubt on this potential is obviously fundamentally damaging to such an account.

The account I endorse, however, requires only that punishments not tend to make reform less likely. Thus it does not link the permissibility of punishments to their ability to bring about positive changes in the offender — it only requires that punishments not get in the way, so to speak, of this prospective outcome. My account is unaffected, therefore, by evidence indicating that impositions of punishment are not effective means of bringing about reform.

The second objection raised against punishments geared toward reform centers not on whether such practices are effective in attaining their stated goals, but rather on whether setting reform as a positive goal of punishment is consistent with the liberal ideal of individual autonomy. Or as it is often expressed in Kantian terms, the objection is that punishments aiming at reform fail to demonstrate proper respect for offenders as persons. Kant's views on punishment remain the subject of substantial debate, but for present purposes the relevant legacy of Kant's account for liberal theories of punishment is that

by punishing in an effort to bring about reform, a political community takes as its goal that which is not rightfully its goal to pursue. Moral reform involves the changing of motives, but moral motives, on the Kantian account, are not the sorts of things appropriately subject to external modification. As Duff characterizes the objection, if a system of punishment aims to reform offenders “by so modifying their dispositions and motives that they will in future willingly refrain from crime, it treats them not as responsible agents who should be left free to determine their own values and attitudes but as objects to be remolded or manipulated into conformity.”³⁸ Instead, if the process of moral reform is to be meaningful, genuine reform, it must be freely chosen by the individual herself. Simply put, reform is not the sort of thing that should be imposed from the outside.

As before, this objection is not damaging to my account. This is, again, because my view does not require that reform be a positive goal of punishment. All that my proposal requires is that punishment not tend to make reform less likely. This requirement is entirely compatible with the view that moral reform must be freely undertaken by the individual herself, as it is also with the recognition that she may never choose to do so.

There is a particular strand of reform-oriented justifications of punishment that attempts to address the previous two objections while still maintaining offender reform as a positive aim. These accounts fall under the broad heading of expressive, or

³⁸ Duff, *Punishment, Communication, and Community*, pp. 13-14.

communicative, theories, according to which punishment is a kind of language; it communicates an important message, namely, that a particular act is condemned by society as being morally wrong. The primary value of this communication may be of various kinds: It may be intrinsically valuable as a vindication of the law or a reaffirmation of the right that was violated³⁹; it may be instrumentally valuable in promoting some societal goal, such as social solidarity⁴⁰ or a morally educated public⁴¹; or, relevant for present purposes, it may be instrumentally valuable in promoting the moral improvement of the wrongdoer herself.

Two prominent varieties of this offender-improvement strand of the communicative account are the moral education theory, according to which the communicative aim of punishment is to teach a wrongdoer that what she did was morally wrong and thus prohibited by society, and Duff's theory of punishment as secular penance, according to which punishment "aims not just to communicate [deserved] censure but thereby to persuade offenders to repentance, self-reform, and reconciliation."⁴² Both of these theories explicitly take offender reform to be a positive aim of punishment; nevertheless, both share two important features with my account. First, both theories accept, as I do, the notion that no one should be treated as morally

³⁹ C.f., Primoratz, "Punishment as Language."

⁴⁰ C.f., Emile Durkheim's account of punishment as a means of fostering solidarity, in his *The Division of Labor in Society*, trans. G. Simpson (New York City: The Free Press, 1964).

⁴¹ C.f., A.C. Ewing, "Punishment as Moral Agency," *Mind* 36 (1927), 297.

⁴² Duff, *Punishment, Communication, and Community*, pp. xvii-xix.

irredeemable.⁴³ Second, and more important for present purposes, communicative theorists emphasize the Kantian ideal of respecting individual autonomy. Thus Jean Hampton, in her seminal development of the moral education theory, writes:

[T]he moral education theorist will admit that the state can predict that many of the criminals it punishes will refuse to accept the moral message it delivers. As I have stressed, the moral education theory rests on the assumption of individual autonomy, and thus an advocate of this theory must not only admit but insist that the choice of whether to listen to the moral message contained in the punishment belongs in the criminal.⁴⁴

On these communicative accounts, then, punishment communicates a message to the offender urging reform, but the offender, being autonomous, may accept or reject the message. By respecting — in fact insisting on, as Hampton writes — the offender’s freedom to reject the message of reform, these communicative theories can answer the two objections leveled against traditional offender-improvement accounts. For the communicative theories, after all, punishment’s aim is not to impose reform on the offender, an aim not only likely to fail but also inappropriate for the penal institutions of a liberal state. Rather, the aim of punishment on these theories is to send a message asking, even urging the offender to choose freely to repent and reform. Given the

⁴³ Duff, for instance, writes that “we can never have *morally* adequate grounds — nothing could count as morally adequate grounds — for treating a person as being beyond redemption.” R. A. Duff, *Trials and Punishments* (Cambridge, U.K.: Cambridge University Press, 1986), p. 266. Similarly, Jean Hampton writes, “the state’s assumption that the people it is entitled to punish are free means it must never regard any one it punishes as hopeless, insofar as it is assuming that each of these persons still has the ability to choose to be moral.” Hampton, “The Moral Education Theory of Punishment,” *Philosophy & Public Affairs* 13 (1984): 208-38, p. 231.

⁴⁴ Hampton, “The Moral Education Theory of Punishment,” p. 230. See also Duff, *Punishment, Communication, and Community*, pp. 80-82, and Falls, “Retribution, Reciprocity, and Respect for Persons,” p. 46.

emphasis such theories place on respecting individual autonomy, as well as their insistence that everyone be treated as having the capacity for redemption, it might seem that I should just endorse one of these accounts, accepting this sort of freely chosen reform as a positive aim of punishment rather than as a prospect not to be undermined.

I actually don't have strong objections to punishments that aim at offender reform, although as I discussed in chapters 1 and 2, I believe the *central* aim of punishment — the reason we should want such an institution — is that it helps to deter criminals and thus to protect the security and well-being of community members. Deterrence, in my view, constitutes both a sufficient and a necessary positive aim of punishment. That is, the role played by the prospect of punishment in reducing crimes constitutes enough of a reason to want such an institution; conversely, if empirical evidence determined that the prospect of punishment had no impact in reducing crime,⁴⁵ and thus that it didn't serve to protect community members security or well-being, then communicating a message of moral censure would not itself constitute a sufficient reason to justify the enormous financial investment that the institution of punishment requires. But if we grant that deterrence is properly the central aim of punishment, I don't object to the sort of moral suasion that Hampton and Duff endorse as a supplementary aim of the practice.

⁴⁵ In fact, the empirical evidence appears to support the intuitive conclusion that criminal sanctions do have a deterrent effect. See, e.g., Daniel S. Nagin, "Criminal Deterrence Research at the Outset of the Twenty-First Century," *Crime and Justice* 23 (1998): 1-42. Nagin writes that "the evidence for a substantial deterrent is much firmer than it was fifteen years ago. I now concur ... that the collective actions of the criminal justice system exert a very substantial deterrent effect" (p. 3). See also Andrew von Hirsch, Anthony E. Bottoms, Elizabeth Burney, and P-O. Wikstrom, *Criminal Deterrence and Sentence Severity: An Analysis of Recent Research* (Oxford, U.K.: Hart Publishing, 1999).

In practice, of course, punishments may tend to have both effects. Also, the distinction between special deterrence (deterring the offender from committing future crimes) and offender reform may in practice become blurrier than is sometimes recognized in theoretical discussions of these aims. Still, if we consider the communicative aspect of punishment and ask who should constitute the primary audience of the institution, in my view the answer is all of us (as potential offenders), not solely those who have already committed crimes (as wrongdoers in need of moral reform). Consider also that if we took the central aim of punishment to be offender reform rather than deterrence, then we would have at least a relatively harder time justifying the punishment of already-repentant offenders, or offenders who may be persuaded to repent by nonpunitive means (nonpunitive censure, perhaps). A drunk driver who kills a person, for instance, may upon sobering up be racked with guilt about what she has done; furthermore, she may genuinely commit to change her behavior, to fight her addiction, and not to make the same mistakes again. Under a system of punishment aimed centrally at reform, it's unclear how we might justify punishing her.⁴⁶ Given deterrence as the institution's central aim, however, it becomes clearer that we would have good reason to punish her in spite of her repentance. The point of punishment is to provide a compelling reason for potential offenders not to commit acts that violate community members' security and well-being. Punishing the already repentant drunk driver would serve to

⁴⁶ C.f., David Boonin, *The Problem of Punishment* (Cambridge, U.K.: Cambridge University Press, 2008), p. 186; and C. L. Ten, "Positive Retributivism," *Social Philosophy and Policy* 7:2 (1990): 194-208, p. 204.

reinforce the credibility of the threat to other potential drunk drivers (many of whom would also undoubtedly be remorseful and repentant if they committed similar acts).

In summary, then, I have no strong objections to the sort of reform endorsed by Hampton and Duff as an aim of punishment, as long as it is not taken to constitute the sole or central aim of the institution. Deterrence constitutes the reason we should want an institution of punishment, and although reform might represent a supplemental aim, the institution's justification does not require that reform be among its aims. On my account, however, the permissibility of punishment does require that penal practices not tend to *undermine* the prospect of offender reform. In the next section, I consider various objections to this thesis.

IV. Objections

First, one might argue that any punishment might undermine the prospect of reform. That is, perhaps any punishment might at least to some degree increase an offender's negative regard for the society on whose behalf she is punished, and thus weaken her motivation to reflect on, and care about, how her criminal actions violated certain of the society's values. If this is the case, then one might think that all punishments, insofar as there exists the possibility that they might undermine reform, would be impermissible. Notice, however, that I have consistently phrased my constraint as a prohibition on punishments that *tend* to undermine the prospect of reform. So for my account to imply a prohibition on all punishments as the objection challenges, it would have to be the case that all modes and degrees of punishment tend to undermine the

prospect of reform. This seems unlikely, unless one believes that impunity always tends to be more conducive to moral reflection and reform than any punishment at all.

This leads to a related point: Undermining the prospect of reform, as I conceive it, is to be distinguished from undermining the mere possibility of reform. The latter formulation could permit a great deal more in terms of punishment, insofar as reform may remain at least a bare *possibility* even for offenders subject to exceedingly harsh or demeaning treatment. Such treatment would undermine the prospect of reform, however, in terms of tending to make the requisite moral reflection less likely than it would have been had punishment not been administered. Thus it would be prohibited on my account.

As the previous two points indicate, which punishments would actually be ruled out by my account will depend significantly on empirical evidence about what tends to undermine the prospect of reform. I do not attempt here to address these empirical questions. Rather, my aim has been to flesh out and argue for the constraining principle itself. But the question of how the principle would constrain in practice — that is, the question of what sorts of punishments actually tend to undermine the prospect of offender reform — is one that I suggest bears further scrutiny. If research found, for instance, that certain forms of punishment significantly correlated with higher frequencies of reoffending (or with offenders' graduating up to more serious crimes) upon release, this would provide at least some tentative evidence that such punishments were undermining

offenders' moral reform.⁴⁷ Again, however, my point in this chapter is not to settle the empirical issue, but rather to defend the principle underlying it.

Finally, one might object that some offenders are simply irredeemable, and if so, we need not concern ourselves with whether punishments might undermine their redemption.⁴⁸ If some criminals are essentially moral monsters, with no potential for reform, then seemingly no punishments would be ruled out by the considerations of reform that I have proposed here. There is no punishment, after all, that will undermine the reform of one who cannot be reformed anyway. We should be careful, however, about ascribing to individuals, even apparently vicious individuals, the labels "moral monster" or "irredeemable." First, it's not clear what would count as adequate evidence that one is truly irredeemable, rather than merely (so far) unredeemed or unrepentant. Even consistently evil behavior is only evidence, in Kantian terms, that one is not currently exhibiting a good will, not that one has no capacity to develop a good will. Kant, in fact, believed that everyone possessed a predisposition to act from respect for the moral law.⁴⁹ This feature of humanity, as Thomas Hill puts it, "implies that anyone who

⁴⁷ A 2007 study by M. Keith Chen and Jesse M. Shapiro, for example, found some evidence that harsher prison conditions led to higher rates of offending after release. See M. Keith Chen and Jesse M. Shapiro, "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach," *American Law and Economics Review* 9:1 (2007): 1-29. Chen and Shapiro's study was limited, however, by a small sample size. Also, a 2004 study by the *Arizona Republic* found that the state's tougher punishment policies for teenage offenders were contributing to nonviolent juvenile offenders' more frequently committing violent crimes upon release. See Judi Villa, "Adult Prisons Harden Teens: Young offenders groomed for life of crime," *The Arizona Republic*, 14 Nov. 2004.

⁴⁸ Rhodes discusses the tendency among officials in maximum security prisons to view prisoners essentially as moral monsters. See Rhodes, *Total Confinement*, esp. pp. 182-90.

⁴⁹ See, e.g., Kant, *Religion within the Boundaries of Mere Reason* 6:28, translated and edited by Allen Wood and George di Giovanni (Cambridge, U.K.: Cambridge University Press, 1998); and *Groundwork of the Metaphysics of Morals* 4:440 (p. 44).

has humanity has a capacity and disposition to follow such principles; but since his rationality may be imperfect or counteracted by other features, he may not always follow these principles.”⁵⁰ Thus Kant believed everyone had the capacity for repentance and reform, that a human being “can never lose entirely his predisposition to the good.”⁵¹

Suppose, however, we are less optimistic than Kant. Suppose we believe that some individuals truly are moral monsters, that some truly have no capacity for reform. Still, there is a significant epistemic challenge in trying to determine which individuals are truly incapable of reform and which are merely as yet unreformed. It’s not clear even how we might go about distinguishing, with any confidence, the former class of individuals from the latter.⁵² Given this practical epistemic challenge, any distinctions we attempt to make between those who can and cannot be redeemed will be suspect. In trying to make such assessments, then, we inevitably risk treating as irredeemable those who are in fact capable of reform. That is, we risk treating some individuals with contempt. Instead, I suggest that even if genuine moral monsters exist, we would do better to err on the side of treating everyone as though she is capable of reform.

⁵⁰ Thomas Hill, Jr., *Dignity and Practical Reason in Kant’s Moral Theory* (Ithaca, N.Y.: Cornell University Press, 1992), p. 41.

⁵¹ *Metaphysics of Morals* 6:463-4 (p. 580).

⁵² Kant cautioned us even about being too confident in distinguishing those who exhibit good wills from those who do not. Given the inscrutability of motives, Kant wrote that “in fact we can never, even by the strictest examination, completely plumb the depths of the secret incentives of our actions” (*Groundwork* 4:407, p. 19). See also Richard Dean, *The Value of Humanity in Kant’s Moral Theory* (Oxford, U.K.: Clarendon Press, 2006). Despite the difficulties of assessing whether a person acts with a good will, however, notice how much more difficult the task would be in determining whether she has no *capacity*, and could not develop the capacity, to do so.

V. Conclusion

In this chapter, I have contended that punishments should not express contempt for offenders, should not give up on them as moral persons. Because of this, punishments should be constrained so that they do not tend to undermine the prospect of offenders' reform. In the previous chapter, I also offered a partial defense of the retributivist principle that punishments should be no more severe than their crimes morally deserve. Both of these constraints essentially set ceilings on how severely we may punish.⁵³ A worry arises, however, with respect to how these constraints fit with what I have claimed is the central aim of punishment, the deterrence of potential criminals. The worry is that for punishments to achieve the goal of deterring potential offenders, these punishments would need to exceed what is allowed by the retributivist and reform-based constraints.⁵⁴

I offer two responses: First, if it did prove to be the case that the floor set by the goal of deterrence were higher than the ceilings set by the retributivist and reform-based constraints, the constraints would take precedence. In other words, the social benefits of punishment are insufficient to justify the institution if its practices fail to respect those punished as moral persons. Second, however, I doubt that the aim of deterrence really does require sentences more severe than the retributivist and reform-based constraints allow. Research into punishment's deterrent effects tends to support the notion that

⁵³ Often the ceiling set by the reform-based constraint will be lower than that set by the retributivist constraint; thus what might be permitted on purely retributivist grounds will often be ruled out insofar as it tends to undermine the prospect of an offenders' reform. In principle, there might also be cases in which the retributivist principle would set the ceiling on punishment lower than the reform-based principle, although such cases seem less likely in practice.

⁵⁴ C.f., Alan Goldman, "The Paradox of Punishment," *Philosophy & Public Affairs* 9:1 (1979): 42-58.

certainty of punishment plays a much larger role than *severity* of punishment. This conclusion undermines the idea that sentences must be especially severe for the institution of punishment to have a significant deterrent impact; effective enforcement appears more relevant to the deterrent impact. Thus it's reasonable, I suggest, to think that the institution of punishment might provide significant deterrent benefits even as it is constrained by the considerations of retribution and reform that I have discussed in these two chapters.⁵⁵

⁵⁵ I appreciate the helpful feedback I received on previous versions of this chapter from Anne Margaret Baxley, Marilyn Friedman, Larry May, Christopher Heath Wellman, and the audience at the 2009 International Social Philosophy Conference.

CHAPTER 5

Collective punishment and distribution of harms

The institutions of international criminal law that have developed in the decades since the Nuremberg trials have consistently focused on prosecuting and punishing individual human agents for their roles in mass crimes. The International Military Tribunal at Nuremberg expressed what has since become the governing view: “Crimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced.”¹ In the wake of atrocities such as the Rwandan and Srebrenica genocides, however, the tribunal’s pronouncement may seem, at least, too simplistic. Genocide and crimes against humanity, after all, are by their nature group endeavors. Individuals may murder, torture, or rape, but no individual alone is responsible for perpetrating mass crimes such as genocide. Rather, such crimes typically result from the actions of groups of individuals, who may be organized in some strong sense or may, at least, influence each other in ways that make possible what would not have been possible from human beings acting individually.

The question of how best to assign responsibility for group wrongdoing has spawned a sizeable philosophical literature.² Scholars have debated whether

¹ International Military Tribunal for the Trial of the Major War Criminals, judgment of Oct. 1, 1946; reprinted in *American Journal of International Law* 41 (1947), p. 221.

² C.f., Larry May, *Sharing Responsibility* (Chicago: University of Chicago Press, 1992), esp. pp. 37-40; see also Larry May and Stacey Hoffman, eds., *Collective Responsibility: Five Decades of Debate in*

responsibility for group-perpetrated crimes is shared, in the sense that responsibility distributes among individual group members, or collective, so that the group itself, as a group, is responsible for the crimes in a nondistributive sense (or whether the best conception of responsibility lies somewhere between these extremes). Because international criminal law views individual human beings as the responsible agents, its challenge has been to prosecute and punish individuals for what they, as individuals, actually did (or failed to do), while nevertheless accounting for the fact that their acts (or omissions) were part of a larger criminal enterprise. Establishing both the criminal act and intent elements for individuals who contributed somehow to the mass crime can be a thorny matter.³

As an alternative, a number of scholars have begun to endorse what I will refer to as “collective punishment,” according to which states themselves, either in addition to or instead of their individual members, are the appropriate subjects of punishment for international crimes. Thus Anthony Lang writes: “[W]hile individuals have been rightfully accused of and punished for these crimes, it seems appropriate that *states* be held responsible as well, for only an organized community has the means to inflict

Theoretical and Applied Ethics (Savage, Md.: Rowman and Littlefield Publishers, Inc., 1991); Peter A. French and Howard K. Wettstein, eds., *Shared Intentions and Collective Responsibility*, spec. issue of *Midwest Studies in Philosophy* XXX (2006); Larry May and Raimo Tuomela, eds., *Collective Responsibility*, spec. issue of *Journal of Social Philosophy* XXXVIII:3 (fall 2007); and Renee Jeffery, ed., *Confronting Evil in International Relations: Ethical Responses to Problems of Moral Agency* (New York City: Palgrave Macmillan, 2008), chapters 1-3.

³ C.f., Larry May, *Crimes Against Humanity: A Normative Account* (Cambridge: Cambridge University Press, 2005), esp. chapters 7-9; and *Aggression and Crimes Against Peace* (New York: Cambridge University Press, 2008), esp. chapter 12; and Mark Drumbl, *Atrocity, Punishment, and International Law* (Cambridge: Cambridge University Press, 2007).

violence on such a large scale.”⁴ On this view, punishing states is often appropriate because states themselves are in many cases the responsible agents of genocide and other mass crimes.⁵ Collective punishment is thus an attractive option insofar as it better ensures that no responsibility for the crime goes unassigned, whereas it has proven difficult to account fully for responsibility for mass crimes by prosecuting and punishing individual contributors.

This paper examines various conceptual and normative questions about how international institutions might implement a practice of collective punishment. Of central concern is whether collective punishment is consistent with treating individual group members with respect as moral agents. Punishment involves inflicting harms on offenders, and the worry with collective punishment is that the harms produced by punishing groups will distribute among group members, some of whom may not have participated in (or may even have worked against) the criminal endeavor. Requiring some group members to bear these distributed burdens even though they did not contribute to the group’s crimes appears inconsistent with respecting them as autonomous moral agents.

In what follows, I consider various strategies of defending collective punishment against the objection that its harms will distribute in unjustifiable ways: First, one might

⁴ Anthony F. Lang Jr., “Crime and Punishment: Holding States Accountable,” *Ethics & International Affairs* 21:2 (Summer 2007): 239-57. See also, e.g., David Luban, “State Criminality and the Ambition of International Criminal Law,” in *Accountability for Collective Wrongdoing*, eds. Tracy Isaacs and Richard Vernon (Cambridge University Press, 2011), pp. 61-91.

⁵ States are not, of course, the only sort of group that may commit international crimes. Nonstate entities, such as terrorist groups, may also commit crimes against humanity or genocide. Although I focus here on collective punishment of states, I note on occasion how the implications might differ for collectively punishing other sorts of group perpetrators.

offer some argument as to why a given distribution of harms is in fact justified. Second, one might accept that collective punishment's harms will distribute unfairly, so that some members of the state will be wronged, but then contend that this presumptive injustice is overridden by the good accomplished (or harm averted) by the practice. Third, one might endorse collective punishment with nondistributive harms. I contend that the first two lines of defense fail. Distributive collective punishment is indeed presumptively unjustified, and the ostensible benefits of such a practice are insufficient to override this presumption. And although collective punishment with nondistributive harms may be morally justifiable, it presents serious difficulties with respect to implementation. Ultimately, international criminal law would do better to maintain its focus on individual human agents, and to continue seeking better ways to fully assign responsibility for mass crimes without imposing punitive burdens on group members who did not contribute.

I. The case for collective punishment

Proponents of collective punishment contend that we should regard the state as the agent of international crimes. Since Hannah Arendt famously wrote of the "banality of evil,"⁶ a number of scholars have pointed out that those who participate in perpetrating mass crimes often aren't properly characterized as criminal deviants, in the sense of deviating from their community's norms; instead, they may be said to act in accordance

⁶ See Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin Books, 1994).

with the norms of a community that is itself criminal.⁷ In circumstances such as these, in which the pathological instead becomes normal, or even expected, it may seem appropriate to hold the group itself, as a group, responsible via collective punishment. Thus David Luban writes: “the Nazi state, like the Hutu Power state half a century later, was in a literal sense criminal to the core. ... A state that turns the world upside down and makes the monstrous the centerpiece of civic obligation is a criminal state.”⁸ If criminality really is pervasive within a group, then collective punishment may more effectively ensure that all of the criminal wrongdoing is accounted for — that no aspect of it goes unpunished.

By ensuring a full assignment of responsibility for mass crimes, collective punishment may be valuable in the message that it sends to citizens of the punished state. Again, given that the factors that allow, or even promote, mass crimes are often pervasive, and sometimes subtle, throughout a group, collective punishment may make it more difficult for some members of the group to avoid accepting responsibility themselves for the roles they played in the crimes. If only the group leaders are punished, then members of the group who escape punishment may find it too easy to rationalize to themselves that they played no role, that they bear no responsibility, for the wrongdoing.

Underlying the idea that collective punishment is a more effective assignment of responsibility for mass crimes is the practical consideration that determining criminal

⁷ C.f., Deirdre Golash, “The Justification of Punishment in the International Context,” in *International Criminal Law and Philosophy*, eds. Larry May and Zachary Hoskins (New York: Cambridge University Press, 2010), p. 212. See also Drumbl, *Atrocity, Punishment, and International Law*, pp. 25-26, 32, 173.

⁸ Luban, “State Criminality and the Ambition of International Criminal Law,” p. 63.

responsibility and punishing mass crimes piecemeal can be extremely difficult. Even if courts could determine each person's individual contributions to the crime and sufficiently establish each person's intent, the process would be exceedingly expensive and time consuming.⁹ Punishing a state as a whole, however, through monetary or other sanctions, appears simpler to accomplish from a practical standpoint.

Finally, even for members of the state who did not contribute to the crimes, collective punishment may provide an incentive to work to oust the criminal leaders, or to change the institutions or culture that allowed the crimes to occur. Philip Pettit writes:

By finding the grouping responsible, we make clear to members as a whole that unless they develop routines for keeping their government ... in check, then they will share in the corporate responsibility of the group; even if they have little or no enactor responsibility, they will have member responsibility for what was done. By finding the grouping responsible in such a case, indeed, we will make clear to the members of other groupings in the same category that they too are liable to be found guilty in parallel cases, should the body to which they belong bring about one or another ill.¹⁰

Thus there appear to be good reasons to give collective punishment serious consideration as an alternative or a supplement to the current system of individually focused prosecutions and punishments. Despite the apparent virtues of collective punishment, however, serious concerns arise regarding how exactly such punishment should be imposed. As suggested before, the central worry is that punishing an entire

⁹ C.f., Avia Pasternak, "Sharing the Costs of Political Injustices," *Politics, Philosophy, and Economics*, republished online November 29, 2010, DOI: 10.1177/1470594X10368260.

¹⁰ Philip Pettit, "Responsibility Incorporated," *Ethics* 117 (2007): 171-201, on p. 200. See also D. J. Levinson, "Collective Sanctions," *Stanford Law Review* 56 (2003): 345-428.

group as a group will nevertheless result in harms to individual members of the group, many of whom may not have participated in the wrongdoing (or may even have actively denounced it or otherwise worked against it). Thus the specter of guilt by association looms. That we may impose harms on an individual in response to a crime in which she played no role, one which she did not even endorse, merely because she is in some sense associated with others who did perpetrate the crime — this notion will strike many as deeply unfair.¹¹ In particular, such treatment appears inconsistent with the principle of respect for persons, which requires at a minimum that our treatment of others be responsive *to them*, to what they have freely done or intended to do.

If collective punishment is defensible in the face of this objection, it appears that the defense will need to proceed along one of the following lines: (a) the harms of collective punishment will distribute, but the distributions are justifiable in themselves; (b) the harms of collective punishment will distribute in presumptively unjustifiable ways, but this presumptive wrong is overridden by some greater good attained through collective punishment; or (c) states can be punished without the harms distributing. I consider each option in turn.

II. Distributing collective punishment's harms

One alternative for an advocate of collective punishment is to concede that its harms will distribute among group members, and then to defend some distribution of the

¹¹ C.f., Toni Erskine, “Kicking Bodies and Damning Souls: The Danger of Harming ‘Innocent’ Individuals While Punishing ‘Delinquent’ States,” *Ethics & International Affairs* 24:3 (2010): 261-285, on pp. 272-73.

harms as justified. An obvious candidate is an equal distribution of harms.¹² But is equal distribution defensible, given the fact that some members of the group will inevitably play a smaller role in the group's crimes than others (or no role, or even work against the crime)?

A number of theorists have argued, on various grounds, that group members, even dissenters, may be liable to bear equally the costs of compensating victims for the harms their group inflicted. David Miller, for instance, argues that in cooperative groups such as democratic political communities, when members reap the benefits of citizenship and have "a fair chance to influence" the community's decisions, then they may be "outcome responsible" for what their group does.¹³ On Miller's view, the equal liability of members of a democratic community stems, in part, from the fact that they participated (or at least had the chance to participate) in the decision-making process. Avia Pasternak, by contrast, rejects appeals to democratic authorization such as Miller's, because she believes community members who actively objected to a harmful policy cannot be said to have authorized it.¹⁴ Instead, Pasternak defends equal distribution of the costs of political injustices by appeal to associative obligations among members of a democratic community.¹⁵

¹² C.f., Avia Pasternak, "The Distributive Effect of Collective Punishment," in *Accountability for Collective Wrongdoing*, eds. Tracy Isaacs and Richard Vernon (Cambridge University Press, 2011), p. 221.

¹³ David Miller, *National Responsibility and Global Justice* (Oxford: Oxford University Press, 2007), p. 119. See also Tony Honoré, *Responsibility and Fault* (Oxford: Hart Publishing, 1999), pp. 78-81.

¹⁴ Pasternak, "Sharing the Costs of Political Injustices."

¹⁵ Ibid.

Rather than engage here with Miller's and Pasternak's contrasting defenses of equally shared consequential responsibility for group wrongs, I want to focus on a point on which their accounts, and a number of other discussions of collective responsibility, agree: the distinction between blameworthiness and consequential responsibility.

Theorists who endorse holding group members, even noncontributors, responsible for their group's wrongdoing are typically quick to clarify that such group members may not be morally blameworthy, or deserving of punishment.¹⁶ Even if it is not permissible, however, to punish individuals for crimes in which they played no role, scholars such as Miller and Pasternak offer various strategies for establishing that group members, even noncontributors or dissenters, may share equally the consequential responsibility for compensating those harmed by the crime. For instance, Miller writes of a polluting, employee-controlled firm whose members decide in a majority vote not to implement more environmentally friendly practices:

It would not in general be right to blame (or punish) members of the minority for what their firm has done to the river — they could quite properly defend themselves by saying that they spoke out against the manufacturing process that caused the pollution. But it is right to hold them, along with others, liable for the damage they have caused.¹⁷

¹⁶ See also, e.g., Anna Stilz, "Collective Responsibility and the State," *The Journal of Political Philosophy*, prepublished online February 12, 2010, DOI: 10.1111/j.1467-9760.2010.00360.x; Iris Marion Young, "Responsibility and Global Labor Justice," *The Journal of Political Philosophy* 12:4 (2004): 365-388; and Hannah Arendt, "Collective responsibility," in *Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt*, ed. James W. Bernauer (Boston: Martinus Nijhoff, 1987).

¹⁷ Miller, *National Responsibility and Global Justice*, p. 119.

Miller's discussion here reflects a common intuition: We typically believe that liability to punishment is an especially serious matter, and so we set a higher standard to hold a person liable to punishment than to hold her liable to pay compensation.¹⁸ There are at least two significant reasons to be especially cautious when punishment is at stake. First, a characteristic feature of punishment is that it expresses blame, or condemnation, of the offender.¹⁹ This feature is not essential to compensation. Thus if a defendant in a civil suit is required to pay compensatory damages, the aim of the decision is not to assess blame but rather to make restitution to those harmed by the defendant's actions. Second, another essential feature of punishments is that they are intended to be of a severity that will harm those punished — whether because offenders deserve the harm, or because the prospect of harm will help deter wrongdoers, etc. It isn't an essential feature, however, of compensation that it be harmful to the agent making compensation. Because of this, impositions of punishment may often be significantly more severe than would be required solely to compensate those harmed. More is typically at stake, then, with punishment than with compensatory liability, both because of punishment's condemnatory message and because particular burdens imposed may often be more severe than compensating victims for their losses would require. These factors may help explain why scholars such as Miller and Pasternak are willing to ascribe consequential responsibility to individual group members where they would not endorse punishment.

¹⁸ This fact is reflected by the greater burden of proof for prosecutors in criminal trials ("beyond a reasonable doubt") than for plaintiffs in civil litigation ("preponderance of the evidence").

¹⁹ See Feinberg, "The Expressive Function of Punishment," in *Doing & Deserving: Essays in the Theory of Responsibility* (Princeton, N.J.: Princeton University Press, 1970).

For present purposes, then, the relevant question is whether collective punishment of a state in which the harms distribute equally among the state's members looks more like a case of assigning consequential responsibility to these members or of punishing them. Pasternak believes it is the former. That is, she contends that equal distribution of the harms of collective punishment does not entail that the punishment itself has distributed; it is still the state itself that is punished, even if its members actually bear the burdens. She justifies her claim by focusing on the expressive or condemnatory aspect of punishment. Because the burdens borne by the group members have no condemning function, she contends, they aren't properly punishment.²⁰ Thus the punishment itself remains at the level of the state; the burdens borne equally by the state's citizens stem from their consequential responsibility to share the burden of compensating those harmed.

There are good reasons, I contend, to believe that the burden distributed among a state's citizens in cases of collective punishment is indeed a punitive burden, rather than merely a compensatory (i.e., consequential) one. First, as indicated above, because punishment and compensation have different aims, the sanction imposed on a perpetrator state as punishment could be significantly more severe than would be required to compensate those harmed by the state's actions. In such cases, the distributed burden shared equally by the state's citizens would similarly be considerably greater than what would be required for compensation. Second, the mode of punishment may differ from what would be required for compensation. Monetary sanctions may be an obvious

²⁰ Pasternak, "The Distributive Effect of Collective Punishment," pp. 216-17. Toni Erskine makes a similar point in "Kicking Bodies and Damning Souls," p. 273.

method of punishing a collectivity such as a state, but advocates of collective punishment have suggested other options. David Luban, for instance, mentions “capital punishment” for the state, that is, “conquest and reconstruction.”²¹ Anthony Lang suggests that punishment could include “coercive military actions, such as the use of aerial bombing against targets as identified as central to the state”²² The harms associated with these punishments, like the burdens of monetary sanctions, would be distributive — they would be borne by the group members. It’s at least unclear, though, how such harms would serve to *compensate* the victims of mass crimes.²³ Thus in their severity and their mode, the burdens shared equally by group members will often be more punitive than compensatory.

Similarly, these considerations cast doubt on whether the condemnatory expression of collective punishment really remains at the level of the state, even as the burdens distribute. If (a) the international community imposes punitive measures on a state, with explicit or implicit recognition that it is asking the state’s citizens to bear the associated burdens, and (b) these burdens are in many cases either more severe or of a different mode than would be appropriate for compensating victims, then (c) it seems reasonable to interpret the message conveyed to the burdened citizens as the message that

²¹ Luban, “State Criminality and the Ambition of International Criminal Law,” p. 90.

²² Lang, “Crime and Punishment,” p. 250. I discuss Lang’s own strategy for defending collective punishment in section III.

²³ However, for an account of how considerations of compensation, or restitution, might be seen to ground a number of sanctions that we typically associate with punishment, see David Boonin, *The Problem of Punishment* (New York: Cambridge University Press, 2008), pp. 231-35. I am skeptical of Boonin’s account, but elaborating on this point is beyond the scope of this paper.

they are being punished for their state's wrongs, not that they are being asked to bear consequential responsibility for their state's harms. This is not to deny that the international community might intend the expression of censure to attach only at the level of the state itself, not to distribute. But it's entirely possible (fairly common, even) to intend to communicate one message but actually to communicate something else. Even if the international community does not intend to communicate censure of the state's citizens themselves, the punitive severity and mode of the harms imposed will tend to do just that. Thus the condemnatory aspect of collective punishment will distribute along with the harms.

For all of these reasons, we would do best to characterize collective punishment in which the harms distribute equally among group members as punishing the group members equally. Given that some group members may have contributed less than others, or not at all, or even worked against the group's crimes, we have good reason to resist this sort of punitive scheme as essentially a form of guilt by association, and thus as inconsistent with basic respect for group members as autonomous moral persons. One can certainly challenge this point, and endorse punishing group members, even those who did not contribute, for their group's crimes (although, again, this tack is explicitly rejected by those such as Miller and Pasternak who want to endorse consequential responsibility). One strategy would be to point to the overriding net benefits to which such a practice would contribute. I discuss this sort of argument in section III.

There are other options for how the harms of collective punishment might be distributed to group members. Pasternak, for instance, also discusses random, or

unpatterned, distribution as well as proportionate distribution.²⁴ First, the harms of collective punishment could distribute among group members according to no real pattern, or at least no pattern imposed by the punishing authority. Consider Luban's "capital punishment" of the state. Such punishment does not lend itself to any clear pattern of harm distribution — equal, proportionate, or otherwise. Even monetary sanctions may not distribute harm in any readily identifiable pattern, if the punished state chooses (and is allowed by the punishing authority) to secure the funds by reducing the services it provides to its citizens. These sorts of unpatterned distributions appear especially unfair, in that they allow for the prospect that those who had nothing to do with, or even worked against, the group's crime may have to bear a greater burden than those who more actively contributed. Furthermore, for the same reasons discussed with respect to equal distribution, we should regard the unpatterned harms borne by group members as genuinely punitive, rather than as consequential (compensatory) burdens. Like equal distribution, then, unpatterned distribution appears unjustified in its own right, although it may be defensible because of its overriding benefits.

Second, collective punishment's harms could be distributed to group members according to their proportionate responsibility. Pasternak notes that such a distribution will be difficult as a practical matter.²⁵ More importantly, though, it's especially unclear that a proportionate distribution scheme could still be said to count as punishment of the group as a group. As I noted above, endorsements of equal distribution schemes rely on

²⁴ Pasternak, "The Distributive Effect of Collective Punishment," pp. 220-28.

²⁵ *Ibid.*, pp. 222-23.

the claim that the harms borne by group members don't amount to punishment, because the condemnatory expression remains at the level of the group itself. I have suggested reasons to be skeptical of this claim, but notice that it is even less plausible in a scheme of collective punishment in which the harms are distributed proportionately. In a proportionate distribution scheme, determinations would need to be made of each group member's relative responsibility, so that the harms could be distributed accordingly. Surely, however, making such determinations is sufficient to change the expressive nature of the punishment. By making determinations of which group member is responsible for what, we in effect shift from censuring the group itself to censuring, to different degrees, the members of the group. So a proportionate distribution scheme, even more than an equal distribution scheme, looks less like merely punishing the group and distributing the harms among members than like punishing the members themselves.

Note, however, that my critique here is not that a proportionate distribution of harms would be unfair in the sense in which I claimed that an equal distribution, and even more so an unpatterned distribution, would be unfair. As I have indicated, I'm ultimately sympathetic to the approach of focusing on individuals, holding them accountable for what they actually did to contribute to the larger endeavor. Rather, my claim is only that a proportionate distribution scheme of collective punishment is not really collective punishment at all. Such a scheme, in practice, would amount to what is the predominant practice in international criminal law today, in which judicial institutions make individual assessments of guilt and punish individual offenders according to what they themselves did to contribute to (or in some cases, did not do to help stop) the group's enterprise.

Ultimately, I doubt that the practice of collective punishment with the harms distributing among group members is intrinsically justifiable. The burden distributed in equal or unpatterned schemes is genuinely a punitive burden, not merely compensatory, and is unfair to those who played no role in the crimes. By contrast, proportional distribution schemes are much fairer, but they no longer count as genuine instances of collective punishment, punishment of the group as a group. In the next section, I consider whether collective punishments whose harms are distributed equally, or randomly, among group members might be all-things-considered permissible, even if presumptively unjustified, because of the net benefits such schemes of punishment produce.

III. Overriding net benefits of collective punishment?

As we've seen, the key move for accounts that attempt to justify some distribution scheme in its own right is to claim that the distributed burdens don't amount to punishment of the group members, to which we typically object for reasons discussed above. Rather, such schemes impose burdens of compensation on group members, and such burdens are somehow justifiable. I have argued that such distribution schemes are best understood as genuine punishment of group members, and that they are not defensible in their own right. But perhaps it is nevertheless all-things-considered permissible to impose punitive harms that will be borne even by those who did not contribute to the group's wrongs, if the net benefits of collective punishment are sufficiently great. Anthony Lang, in endorsing regime change and lustration as state

punishment, admits that “this punishment might result in the harm or death of individuals who had nothing to do with the policies of the government.”²⁶

As with any use of force in the international system, avoiding such deaths should be a primary goal. Nonetheless, concern with such deaths cannot prevent the use of punitive measures to enforce norms. Indeed, one might argue that the seriousness of genocide demands that punitive measures be employed even knowing there is a significant chance that innocents might die, and regrettable as that is.²⁷

For Lang, collective punishment might be justified (or required), even if some of the group members punished contributed nothing to the group’s crimes, if such a scheme were an effective way of preventing atrocities such as genocide. It’s worth considering, then, whether the net benefits of collective punishment would be sufficient to render presumptively unjustified punishment of group members all-things-considered justified. There are, I suggest, reasons to be skeptical of this claim.

In evaluating Lang’s claims about the net benefits of collective punishment, it’s important to bear in mind the distinction between, on one hand, the use of force against some perpetrator state as a means of warding off its attack against some victim group and, on the other hand, the use of force to punish a perpetrator state. Typically, self-defense or defense of others is understood as the imposition of force on perpetrators to stop them from committing (or continuing to commit) some wrongful act. Punishment, by contrast, imposes harm on offenders not to ward off their impending or ongoing attacks, but rather

²⁶ Lang, “Crime and Punishment,” p. 255.

²⁷ Ibid. Also see p. 253.

in response to some wrongdoing they have already committed.²⁸ Lang appears to conflate these two distinct notions — stopping crimes that are being committed and punishing crimes that have been committed — when he writes that one effect of punishing states may be “retributive, in that the specific state committing the crime would be forced to stop.”²⁹ But we cannot defend collective punishment, including the regrettable killing of innocent group members, by appeal to the overriding benefits of the use of international force to stop some ongoing genocide. Rather, collective punishment must be defensible as an imposition of harm on some state for crimes it has already committed.

As discussed earlier, one commonly cited rationale for holding groups collectively responsible is that in many cases the group itself is best understood as the criminal agent, and thus a full assignment of responsibility requires that the group itself be held accountable. Philip Pettit contends that assigning responsibility solely among those individuals who enact the group’s plan “may leave a deficit in the accounting books, and the only possible way to guard against this may be to allow for the corporate responsibility of the group in the name of which they act.”³⁰ Assigning responsibility

²⁸ If the aim of punishment is to deter potential wrongdoers from committing future offenses, then punishment may appear more closely analogous to self-defense. Indeed, a number of authors have attempted to justify deterrent punishment by appeal to considerations of self-defense. Nevertheless, inflicting harm in self-defense and inflicting harm as deterrent punishment are distinct notions in need of distinct justifications. The former involves using force against some attacker who is actually threatening or inflicting harm in order to prevent or stop the attack. The latter involves harming an offender in response to some already-committed act, with the aim of discouraging the offender or other potential offenders from acting similarly in the future.

²⁹ Lang, “Crime and Punishment,” p. 253. Lang’s claim is doubly confusing given that, insofar as forcing a state to stop committing a crime resembles any sort of punishment, it resembles deterrent punishment, not retributive punishment.

³⁰ Philip Pettit, “Responsibility Incorporated,” *Ethics* 117 (2007): 171-201, p. 194. See also Lang, “Crime and Punishment,” pp. 243-46; Luban, “State Criminality and the Ambition of International Criminal Law,”

solely among individual group members may leave this “deficit in the accounting books” because the group’s institutional structures and the relations among group members may play some causal role in the crime as well. Holding the group itself accountable may better ensure a full assignment of responsibility for the crime.

Ensuring this sort of complete accounting might be regarded as valuable in its own right, especially by those who take retributivism to be the goal of an institution of punishment. In my view, however, the central aim of the practice of punishment is deterrence rather than retribution.³¹ Thus the relevant question is whether a system of collective punishment would sufficiently promote the aim of preventing future perpetration of international crimes so as to override the presumptive injustice of imposing punishment on citizens who did not participate in the state’s crimes.

First, we might consider the relative deterrent impact of collective punishment on those group leaders who develop and initiate the criminal plan. In these cases, it’s doubtful that such a punishment scheme would be more effective as a deterrent than individual prosecutions and punishments. Presumably, these leaders would prefer collective punishment in which the harms are distributed among community members.

pp. 78-82; Kirsten Ainley, “Individual Agency and Responsibility for Atrocity,” and Arne Johan Vetlesen, “Collective Evildoing,” in Jeffery, *Confronting Evil in International Relations*.

³¹ Many contemporary scholars argue instead that the central aim of punishing international crimes is to express condemnation of the crime to the perpetrators, the victims, or to the international community generally. See, e.g., Drumbl, *Atrocity, Punishment, and International Law*. In my view, however, although there is a genuine expressive aspect to punishment, the relevant question for accounts that appeal to this aspect is why it is sufficiently important to *justify* punishment. Typically, expressive accounts will appeal to the role such condemnations can play in preventing future atrocities or to the intrinsic value of expressing deserved censure of wrongdoers. Ultimately, then, I believe the expressivist justifications of punishment are best characterized either as deterrent or retributive in their aims.

Given the disproportionate role the leaders play in perpetrating the crime, they would have good reason to think they'd fare comparatively better in a collective punishment scheme — in which harms are distributed equally, or randomly, etc. — than if everyone were punished individually in proportion to what she actually did.³²

Perhaps, instead, the threat of collective punishment would be more effective than individual punishment at deterring potential minor participants, those who might not develop or initiate the plan but nevertheless might play some active, culpable role in the criminal endeavor. Or maybe the threat of collective punishment would more effectively deter those who might otherwise be complicit in the crime — that is, those who, though they are not the principal planners or actors, nevertheless through their actions or failures to act might make the crime more likely to occur.³³ In one sense, this appears more plausible than in the case of leaders, given the greater likelihood that minor participants or those who aid or abet the principal agents might slip through the net of individual prosecutions and punishments, whereas they would be liable to share the punitive harms of collective punishment.

Suppose it's true that the threat of collective punishment would better discourage potential minor players than would the existing system of individual prosecutions and punishments. We might take this either as a point in favor of collective punishment or as

³² It's true that the prospect of collective punishment in addition to (rather than instead of) traditional individual prosecutions and punishment might yield some additional deterrent effect, although this isn't a given. I say more below about collective punishment as augmenting rather than replacing the traditional scheme of individual punishments.

³³ See, e.g., Judgment, *Akayesu* (ICTR-96-4-T), Trial Chamber, 2 September 1998, para. 537. For a useful recent discussion of legal, and moral, complicity, see Larry May, "Complicity and the Rwandan Genocide," *Res Publica* 16 (2010): 135-152.

a reason to continue to develop and improve the existing system. In other words, as we ask whether collective punishment would yield sufficiently greater deterrent effects to override the presumption against it, the appropriate comparison class is not necessarily the current system of individual prosecutions and punishments, but rather the best realistically attainable version of such a system. Under the current system, prosecuting and punishing large numbers of the minor players or accomplices in mass atrocities may be so expensive or time-consuming as to be a practical impossibility. But there might be ways to develop and improve the current system, to address these practical issues.³⁴ If so, then the ostensible comparative advantage of collective punishment in deterring minor contributors would be diminished, or even negated.

In fact, however, there is reason to doubt that the threat of collective punishment would be more effective as a deterrent of potential minor players than would a scheme of individual prosecutions and punishments — even an imperfect scheme in which some contributing parties slipped through the cracks. Consider which would more likely deter a potential minor participant: (a) the prospect that, if the crime is prosecuted and punished, each group member will share the punitive harms regardless of her involvement, or lack thereof, in the criminal endeavor; or (b) the prospect that, if the crime is prosecuted and

³⁴ For instance, a number of scholars have expressed tentative optimism about the *gacaca* courts being implemented in the wake of the Rwandan genocide. *Gacaca* are local courts; suspected participants and accomplices are tried in the villages where they allegedly committed their crimes, but with oversight and procedures aimed at ensuring due process. Although the fairness of *gacaca* courts has been disputed, they have, as a practical matter, allowed for many more trials of low-level participants and accomplices than could have been conducted by the International Criminal Tribunal for Rwanda or the International Criminal Court. C.f., Drumbl, *Atrocity, Punishment, and International Law*, and May, “Complicity and the Rwandan Genocide.”

punished, individual determinations will be made based on what role each person played in the larger endeavor (as planner, enactor, accomplice, etc.). Assuming she is not herself capable of stopping the crime, then if the crime is successfully prosecuted, (a) threatens punitive harms regardless of whether she participates or not; (b), on the other hand, represents the opportunity to escape eventual punishment by not contributing. Thus whereas (b) offers her a clear reason not to participate, or aid or abet those who do participate, (a) offers her no real reason *not* to contribute to the enterprise. In a collective punishment scheme, if the crime is not successfully prosecuted, a group member escapes punishment whether or not she contributes; if the crime is successfully prosecuted, she shares in collective punishment's harms to the same degree whether or not she contributes. Thus, in this sense, collective punishment appears to represent a less compelling deterrent threat for lesser participants and accomplices than does a scheme of individual prosecution and punishment.

One might object that if enough citizens either did not participate or actively worked to prevent their state's crime (by undermining the government's efforts, or even ousting the leadership), the crime might be averted. And the threat of collective punishment might better deter this way, by motivating dissent among the state's membership, than would the threat of individual punishments.³⁵ Similarly, even if the

³⁵ Philip Pettit writes: "By finding the grouping responsible, we make clear to members as a whole that unless they develop routines for keeping their government or episcopacy in check, then they will share in the corporate responsibility of the group; even if they have little or no enactor responsibility, they will have member responsibility for what was done. By finding the grouping responsible in such a case, indeed, we will make clear to the members of other groupings in the same category that they too are liable to be found guilty in parallel cases, should the body to which they belong bring about one or another ill." Pettit, "Responsibility Incorporated," p. 200.

state carried out the crime, the imposition of collective punishment might be sufficiently burdensome for citizens that they would be motivated to work to prevent such crimes in the future (again, perhaps by undermining the leadership's efforts or overthrowing it entirely). Notice, however, that the relative deterrent advantage of collective punishment in these cases depends on the citizens' each believing that enough other citizens will act the same way that their efforts to prevent the crime will succeed. By contrast, if a citizen believes the crime will or won't occur regardless of what she does, then as the previous discussion indicated, the threat of collective punishment will give her less reason not to contribute than would the threat of individually determined punishments. Also, among those opposed to or at least ambivalent toward their government's activities, the imposition of collective punishment could foster resentment of the punishing entity for harming indiscriminately, and perhaps even increased solidarity with their own leaders. The decades-long U.S. trade embargo against Cuba, for instance, has by many accounts tended to increase the Cuban people's solidarity with the Castro regime. For all of these reasons, then, I am skeptical of the claim that the threat or imposition of collective punishment would foment more resistance to perpetrator regimes than would individual punishments to a sufficient degree to override the presumption against collective punishment.

The considerations discussed here are not, I recognize, conclusive, but they need not be. We began with a presumption against collective punishment, because it imposes harms on individuals for what others have done — thus it fails to respect them as autonomous moral agents. If one is nevertheless to endorse collective punishment

because of the overriding net benefits it yields, the burden falls to the advocate to present a compelling case for these benefits. Unless this can be done, collective punishment will be an inappropriate alternative to individual punishment.

Perhaps, though, even if the benefits of collective punishment alone are insufficient to override the presumption against it, we should still endorse collective punishment in addition to (rather than instead of) individual punishment.³⁶ Now, however, we have the opposite problem from that with which we started. As discussed earlier, those who support collective punishment often cite what Pettit calls “a deficit in the accounting books” left from individual punishments alone. By punishing the entire group, as a group, we ensure that all group members are held accountable (even, unfortunately, those who did not contribute to the crime). But if we layer collective punishment on top of individual punishments, there’s a worry that we may instead be left with a *surplus* in the accounting books. At least, we need to be clear about what exactly is the crime for which we’re punishing the group, and what are the crimes for which we’re punishing the group members. The relationship between the group’s crime and the group members’ crimes shouldn’t be that of a whole to its parts, or this would amount to an overassignment of criminal responsibility. And because the collective punishment’s harms would distribute (equally, randomly, etc.) among group members, there’s a significant likelihood that some group members (those who have been prosecuted and punished as individuals) would be doubly punished — that is, they would bear punitive

³⁶ See, e.g., Lang, “Crime and Punishment,” pp. 239-40; Pettit, “Responsibility Incorporated,” pp. 192-98.

burdens for their own contributions to the crime, and they would bear additional punitive burdens as members of the group that committed the crime. It would be essentially equivalent to punishing a three-person team of bank robbers for their recent robbery, and also punishing the first member for helping plan the crime and driving the getaway car, the second member for helping plan the crime and holding the gun, and the third member for helping plan the crime and issuing the demand.

Note that my concern here is not, strictly speaking, with the severity of the harms imposed. It would seem strange indeed to worry that a leader who developed and initiated a genocidal plan resulting in hundreds of thousands of deaths might not only, say, receive a sentence of life in prison but might also be forced to pay an equal share of some monetary sanction imposed on the group. My objection, rather, is with requiring that he suffer whatever harm is deemed sufficient for his contribution to the group's crime, and then also requiring that he suffer additional harm for being a member of the group that committed the crime. In section II, I discussed the worry that collective punishment schemes would impose genuine punishment (not merely compensatory burdens) on innocent group members. Here, the additional worry is that such a punishment scheme, in tandem with individual punishments, would in effect punish guilty group members twice in response to the same crime.³⁷ Thus the presumption against such a punishment scheme

³⁷ In fact, something like this occurs at the domestic level in U.S. criminal law. Douglas Husak describes the practice of "charge stacking," whereby prosecutors may "bring a number of charges against a defendant for the same underlying conduct." Douglas Husak, *Overcriminalization: The Limits of the Criminal Law* (New York City: Oxford University Press, 2008), p. 22. Husak explains, "As long as these offenses contain distinct elements, no rule or doctrine automatically prevents the state from bringing several charges simultaneously, even though, from the intuitive perspective of a layperson, the defendant has committed

should be even stronger than before, and again, it's not apparent that such a scheme would yield sufficiently greater deterrent effects, compared with the best feasible system of individual prosecutions and punishments, to override this presumption.

Rather than punishing the group (and by distribution, group members) for committing the crime and then punishing group members individually for their roles in committing the crime, we might endorse punishing citizens for their individual contributions to the crimes and then punishing the group collectively for allowing itself to be structured and organized in such a way as to allow the crimes to occur.³⁸ Thus the relationship of the state's crime and the citizens' crimes wouldn't be that of a whole to its parts. This appears more promising, in that it appears to avoid unfairly punishing some members of the state twice for the same crime. Still, I contend that such a scheme would be ultimately unfair to those citizens who are not responsible for helping to create the institutional, cultural, or other factors that facilitated the state's criminal acts. The unfairness of imposing burdens on innocent citizens for the crimes of others sets the presumptive bar against collective punishment quite high, I believe, and it's just not clear that the added deterrent impact of such a scheme, whether instead of or in addition to individual punishments, is sufficient to overcome this presumption.

but a single crime." In my view, imposing multiple punishments on an individual for the same criminal act is unjustified whether at the domestic or international level.

³⁸ C.f., Pettit, "Responsibility Incorporated," p. 197.

One way to avoid this worry about overpunishing some group members would be to devise a scheme according to which collective punishment's harms would not distribute. I turn now to consider the possibility of nondistributive collective punishment.

IV. Nondistributive collective punishment

Suppose individual group members could be punished for their contributions to a crime, and also the collective itself could be punished, but the collective punishment's harms would not distribute among the group members. Such a scheme would have the merit not only of avoiding overpunishment of guilty group members, but punishment of innocent group members as well. In addition, it would presumably satisfy advocates of collective punishment who contend that such punishment is necessary to ensure a full accounting for the mass atrocity.

But how might nondistributive collective punishment work? How could groups be punished in such a way that the punitive harms would not trickle down to members of the group? Certainly monetary sanctions would impact group members, as would embargoes, boycotts, or the targeted military strikes such as Lang endorses. It's not immediately clear, then, how a punitive burden could be imposed on the group itself with this burden ultimately being distributed among some number of group members.³⁹

Advocates of holding groups collectively responsible typically contend that the actions of the group itself are often not entirely reducible to the contributions of the

³⁹ C.f., Erskine, "Kicking Bodies and Damning Souls," esp. pp. 274-79.

individual leaders, participants, accomplices, etc.⁴⁰ On one version of this claim, the corporate structure of the group itself, the institutional decision-making procedures, distribution of authority, etc., play a causal role in the perpetration of the crime that cannot be accounted for among the contributions of the individual group members. If so, then we might think that nondistributive collective punishment could target these institutional aspects of the group, replacing the problematic procedures and structures with more appropriate ones. This might be a way of inflicting harm on the group itself, by encroaching on its sovereignty and altering its structure, and also expressing condemnation of it. Thus it might be seen as genuine punishment of the group itself. The question is whether the harms associated with such punishment would ultimately distribute among group members.

On one hand, if the newly imposed institutions were more effective, less susceptible to corruption, etc., than those they replaced, this might on average benefit group members. On the other hand, it's hard to imagine how such institutional restructuring could be accomplished without some individuals being harmed, most obviously through loss of their jobs. To the extent that those most likely to be harmed by the institutional restructuring may be the group leaders and other active participants in the international crime, we may not be particularly troubled by this prospect. Regardless of which individuals bear the burdens of punitive institutional restructuring, however, so long as individuals do bear these burdens, this will not count as a case of *nondistributive*

⁴⁰ C.f., Pettit, "Responsibility Incorporated," pp. 194-98; and D. E. Cooper, "Collective Responsibility," in May and Hoffman, eds., *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*, pp. 35-46, esp. pp. 37-38.

collective punishment. Rather, it is an instance of distributive collective punishment, and as such it is subject to the concerns I raised in sections II and III. In particular, if those harmed by punitive restructuring are the leaders and other active participants, the specter of double punishment again arises, given that what's at issue here is the prospect of collective punishment alongside individual punishments.

There might, of course, be other ways to punish collectively with nondistributive harms. Perhaps, for instance, the international community might take away some of a state's territory. If the territory in question was populated, however, questions would arise about whether this would represent a harm to those residents. Even if the territory was unpopulated but contained significant natural resources, questions of harm might arise with respect to loss of use of those resources.

Ultimately, I'm skeptical about whether a state can be harmed without the harm distributing to any of its citizens. If the international community imposes some burden on a state, the state itself cannot be the bearer of that burden independently of its members. Even if the state in fact consists in more than just the aggregation of its members — even if, e.g., it also consists in the institutional structures that govern its members' interactions — these institutional structures themselves aren't the kind of things that can be harmed, because they aren't the kind of things that have interests. It appears, then, that for any sort of collective punishment we consider, either (a) it will harm some members of the state, or (b) it will not harm at all. If (a), then it will be distributive collective punishment, and as such it will face the objections raised in sections II and III. If (b), then it will not properly constitute punishment, given that the imposition of harm is an essential feature

of punishment. In the end, then, it looks as though the notion of collective punishment with nondistributive harms is not only practically intractable, but perhaps conceptually mistaken.

VI. Conclusion

Despite the appeal of collective punishment as a means of ensuring that full responsibility is assigned for international crimes, I have argued that difficulties arise in how such a punishment scheme could be implemented. Given that distributing the harms of punishment among group members ultimately amounts to punishing group members, such an approach will almost certainly result in the harming of those who played no role in (or even worked against) the crime. And if the distributive collective punishment scheme operates alongside a system of individual punishments, there will be the additional hazard of subjecting some group members (i.e., leaders and participants) to double punishment. Either way, such a punishment scheme appears presumptively unjust, and it's doubtful whether the ostensible gain in deterrent impact from such a scheme would be sufficient to override this presumption. By contrast, nondistributive collective punishment appears, in principle, consistent with respecting individual group members as moral persons. In practice, however, it's not clear how we might punish the group itself without the punitive burdens being borne by its members.

It's worth emphasizing, in conclusion, that although I oppose collective punishment, I recognize the imperfections of the current system of international criminal law. Existing legal institutions don't have sufficient financial or human resources, or time,

to prosecute and punish all those who, through their participation or their complicity, contribute to genocides and other mass crimes. And as a practical reality, even with modifications and improvements, a system of individually focused punishments may never be able to assign full responsibility for such crimes. Some responsibility may inevitably slip through the cracks, either in the form of structural or institutional forces that facilitated the crimes, or individual citizens whose contributions, though real, are small enough to escape attention. Nevertheless, given the significance of what's at stake with punishment, and the strong presumption in favor of respecting people as autonomous moral beings — which involves treating them according to their own voluntary acts and omissions — I suggest the international community should continue to focus on individual criminal prosecutions and punishments rather than collective punishment.⁴¹

⁴¹ I am grateful to Larry May, Christopher Heath Wellman, David Wood for helpful comments on an earlier version of this paper. Also, I presented a draft of the paper in October 2010 at a departmental dissertation workshop at Washington University in St. Louis. I appreciate all the thoughtful questions and suggestions from the participants at that session, including Nate Adams, Jill Delston, John Gabriel, Jason Gardner, Don Goodman-Wilson, Jan Plate, and Bryan Stagner.

WORKS CITED

- Ainley, Kirsten. (2008) "Individual Agency and Responsibility for Atrocity." In Renee Jeffery, ed. *Confronting Evil in International Relations: Ethical Responses to Problems of Moral Agency*. New York City. Palgrave Macmillan.
- Akayesu judgment. (1998) ICTR-96-4-T. Trial Chamber.
- Arendt, Hannah. (1994) *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York City. Penguin Books.
- Arendt, Hannah. (1987) "Collective responsibility." In James W. Bernauer, ed. *Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt*. Boston. Martinus Nijoff.
- Armstrong, K. G. (1969) "The Retributivist Hits Back." In *The Philosophy of Punishment*, ed. H. B. Acton. London: Macmillan.
- Ashworth, Andrew. (1989) "Criminal Justice and Deserved Sentences." *Criminal Law Review*, 340-55.
- Bedau, Hugo Adam. (1978) "Retribution and the Theory of Punishment." *The Journal of Philosophy* 75(11), 601-20.
- Benn, S. I. (1958) "An Approach to the Problems of Punishment." *Philosophy* 33(127), 325-41.
- Bentham, Jeremy. (1996) *An Introduction to the Principles of Morals and Legislation*. Ed. J. H. Burns and H. L. A. Hart. Oxford: Clarendon Press.
- Blumstein, A., Cohen, J., and Nagin, D., eds. (1978) *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, D.C.: National Academy of Sciences.
- Boonin, David. (2008) *The Problem of Punishment*. New York City: Cambridge University Press.
- Braithwaite, John, and Pettit, Philip. (1990) *Not Just Deserts: A Republican Theory of Criminal Justice*. Oxford: Clarendon Press.
- Burgh, Richard. (1982) "Do the Guilty Deserve Punishment?" *The Journal of Philosophy* 79(4), 193-210.

- Carrier, Fanny. (2008, June 25) "Top U.S. court slashes Exxon Valdez oil damages." *Agence France Presse*.
- Chen, M. Keith, and Shapiro, Jesse M. (2007) "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach." *American Law and Economics Review* 9(1), 1-29.
- Cooper, D. E. (1991) "Collective Responsibility." In Larry May and Stacey Hoffman, eds. *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*. Savage, Md. Rowman and Littlefield Publishers, Inc.
- Corlett, J. Angelo. (2001) "Making Sense of Retributivism." *Philosophy* 76(295), 77-110.
- Cottingham, John. (1979) "Varieties of Retribution." *The Philosophical Quarterly* 29, 238-46.
- Dagger, Richard. (2008) "Punishment as Fair Play." *Res Publica* 14, 259-75.
- Dagger, Richard. (1993) "Playing Fair with Punishment." *Ethics* 103, 473-88.
- Darwall, Stephen. (1977) "Two Kinds of Respect." *Ethics* 88(1), 36-49.
- Davis, Michael. (1993) "Criminal Desert and Unfair Advantage: What's the Connection?" *Law and Philosophy* 12, 133-56.
- Davis, Michael. (1992) "Using the Market to Measure Deserved Punishment: A Final Defense." In Michael Davis, ed. *To Make the Punishment Fit the Crime: Essays in the Theory of Criminal Justice*. Boulder, Colo.: Westview.
- Davis, Michael. (1983) "How to Make the Punishment Fit the Crime." *Ethics* 93, 726-52.
- Dean, Richard. (2006) *The Value of Humanity in Kant's Moral Theory*. Oxford: Clarendon Press.
- Dolinko, David. (1991) "Some Thoughts about Retributivism." *Ethics* 101(3), 537-59.
- Drumbl, Mark. (2007) *Atrocity, Punishment, and International Law*. Cambridge, U.K. Cambridge University Press.
- Duff, R. A. (2008) "Legal Punishment." *Stanford Encyclopedia of Philosophy*. Online at <http://plato.stanford.edu/entries/legal-punishment/> (accessed Sept. 8, 2008).

Duff, R. A. (2001) *Punishment, Communication, and Community*. Oxford, U.K.: Oxford University Press.

Duff, R. A. (1986) *Trials and Punishments*. Cambridge, U.K.: Cambridge University Press.

Duff, R. A., and Garland, David, eds. (1994) *A Reader on Punishment*. Oxford: Oxford University Press.

Durkheim, Emile. (1964) *The Division of Labor in Society*. Trans. G. Simpson. New York City: The Free Press.

Erskine, Toni. "Kicking Bodies and Damning Souls: The Danger of Harming 'Innocent' Individuals While Punishing 'Delinquent' States." *Ethics & International Affairs* 24: 261-285.

Ewing, A.C. (1927) "Punishment as Moral Agency: An Attempt to Reconcile the Retributive and the Utilitarian View." *Mind* 36 (143), 297-305.

Falls, M. Margaret. (1987) "Retribution, Reciprocity, and Respect for Persons." *Law and Philosophy* 6(1), 25-51.

Feinberg, Joel. (1970) "The Expressive Function of Punishment." In *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton, N.J.: Princeton University Press.

French, Peter A., and Wettstein, Howard K., eds. (2006) *Shared Intentions and Collective Responsibility*. Special issue of *Midwest Studies in Philosophy* XXX.

"Gently Does It; Prison Conditions." (2007, July 28) *The Economist*, U.S. edition.

Gladwell, Malcolm. (2005) *Blink: The Power of Thinking Without Thinking*. New York City: Little, Brown and Company.

Golash, Deirdre. (2010) "The Justification of Punishment in the International Context." In Larry May and Zachary Hoskins, eds. *International Criminal Law and Philosophy*. New York City. Cambridge University Press.

Golash, Deirdre. (2005) *The Case Against Punishment: Retribution, Crime Prevention, and the Law*. New York City and London: New York University Press.

Goldman, Alan. (1979) "The Paradox of Punishment." *Philosophy & Public Affairs* 9(1), 42-58.

- Gottman, John M.; Murray, James D.; Swanson, Catherine C.; Tyson, Rebecca; and Swanson, Kristin R. (2002) *The Dynamics of Marriage: Dynamic Linear Models*. Cambridge, Mass.: MIT Press.
- Hampton, Jean. (2007) *The Intrinsic Worth of Persons: Contractarianism in Moral and Political Philosophy*, ed. Daniel Farnham. New York City: Cambridge University Press.
- Hampton, Jean. (1991) "A New Theory of Retribution." In C. Morris and R. Frey, eds. *Liability and Responsibility*. Cambridge: Cambridge University Press.
- Hampton, Jean. (1984) "The Moral Education Theory of Punishment." *Philosophy & Public Affairs* 13, 208-38.
- Hanna, Nathan. (2008) "Say What? A Critique of Expressive Retributivism." *Law and Philosophy* 27, 123-50.
- Hart, H. L. A. (1994) *The Concept of Law* 2d ed. Oxford: Oxford University Press.
- Hart, H. L. A. (1968) *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Oxford University Press.
- Hart, H. L. A. (1955) "Are There Any Natural Rights?" *The Philosophical Review* 64(4), 175-91.
- Hegel, G. W. F. (1942) *The Philosophy of Right*. Trans. T. Knox. Oxford: Oxford University Press.
- Hill, Jr., Thomas E. (2000) *Respect, Pluralism, and Justice: Kantian Perspectives*. Oxford: Oxford University Press.
- Hill, Jr., Thomas E. (1992) *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, N.Y.: Cornell University Press.
- Honderich, Ted. (2006) *Punishment: The Supposed Justifications Revisited*. London: Pluto Press.
- Honoré, Tony. (1999) *Responsibility and Fault*. Oxford, U.K. Hart Publishing.
- Hospers, John. (1977) "Punishment, Protection, and Retaliation." In J. B. Cederblom and William L. Blizek, eds. *Justice and Punishment*. Cambridge, Mass.: Ballinger Publishing Co.

International Military Tribunal for the Trial of the Major War Criminals. (1947) Judgment of Oct. 1, 1946. Reprinted in *American Journal of International Law* 41.

Jeffery, Renee, ed. (2008) *Confronting Evil in International Relations: Ethical Responses to Problems of Moral Agency*. New York City. Palgrave Macmillan.

Kalderon, Mark. (2005) *Moral Fictionalism*. Oxford: Oxford University Press.

Kant, Immanuel. (1998) *Religion within the Boundaries of Mere Reason*. Trans. and ed. Allen Wood and George di Giovanni. Cambridge, U.K.: Cambridge University Press.

Kant, Immanuel. (1997) Georg Ludwig Collins' and Johann Vigilantius' notes from Kant's ethics lectures. In Peter Heath, trans., and Peter Heath and J.B. Schneewind, eds. *The Cambridge Edition of the Works of Immanuel Kant, Lectures on Ethics*. Cambridge, U.K.: Cambridge University Press.

Kant, Immanuel. (1996) *The Metaphysics of Morals*. In Mary J. Gregor, trans. and ed. *The Cambridge Edition of the Works of Immanuel Kant, Practical Philosophy*. Cambridge, U.K.: Cambridge University Press.

Kant, Immanuel. (1994) *Groundwork of the Metaphysics of Morals*. In James W. Ellington, trans. Immanuel Kant, *Ethical Philosophy*, 2d ed. Indianapolis, Ind.: Hackett Publishing, Inc.

Kelsen, Hans. (1946) *General Theory of Law and State*. Cambridge, Mass.: Harvard University Press.

Klosko, George. (2005) *Political Obligations*. Oxford: Oxford University Press.

Klosko, George. (1992) *The Principle of Fairness and Political Obligation*. Lanham, Md.: Rowman & Littlefield Publishers, Inc.

Lacey, Nicola (1988) *State Punishment: Political Principles and Community Values*. London: Routledge.

Lang Jr., Anthony F. (2007) "Crime and Punishment: Holding States Accountable," *Ethics & International Affairs* 21: 239-57.

Levinson, D. J. (2003) "Collective Sanctions." *Stanford Law Review* 56: 345-428.

Locke, John. (1996) *Second Treatise of Civil Government*. In Robert M. Stewart, ed. *Readings in Social & Political Philosophy*. New York City: Oxford University Press.

- Luban, David. (2011) "State Criminality and the Ambition of International Criminal Law." In Tracy Isaacs and Richard Vernon, eds. *Accountability for Collective Wrongdoing*. Cambridge University Press.
- Mackie, J. L. (1982) "Morality and the Retributive Emotions." *Criminal Justice Ethics*, 3-10.
- Mason, Michelle. (2003) "Contempt as a Moral Attitude." *Ethics* 113, 234-272.
- Matravers, Matt. (2000) *Justice and Punishment*. Oxford: Oxford University Press.
- May, Larry. (2010) "Complicity and the Rwandan Genocide." *Res Publica* 16: 135-152.
- May, Larry. (2008) *Aggression and Crimes Against Peace*. New York City. Cambridge University Press.
- May, Larry. (2005) *Crimes Against Humanity: A Normative Account*. Cambridge, U.K. Cambridge University Press.
- May, Larry. (1992) *Sharing Responsibility*. Chicago. University of Chicago Press.
- May, Larry, and Hoffman, Stacy, eds. (1991) *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*. Savage, Md. Rowman and Littlefield Publishers, Inc.
- May, Larry, and Tuomela, Raimo, eds. (2007) *Collective Responsibility*. Special issue of *Journal of Social Philosophy* XXXVIII:3.
- McCloskey, H. J. (1967) "Utilitarian and Retributive Punishment." *Journal of Philosophy* 64(3), 91-110.
- McCloskey, H. J. (1957) "An Examination of Restricted Utilitarianism." *The Philosophical Review* 66(4), 466-85.
- Mill, John Stuart. (2001) *Utilitarianism*, 2d ed., ed. George Sher. Indianapolis, Ind.: Hackett Publishing.
- Miller, David. (2007) *National Responsibility and Global Justice*. Oxford. Oxford University Press.
- Miller, William Ian. (1997) *The Anatomy of Disgust*. Cambridge, Mass.: Harvard University Press.

- Morris, Herbert (1968) "Persons and Punishment." *Monist* 52, 475-501.
- Murphy, Jeffrie G. (1978) *Retribution, Justice, and Therapy*. Boston: D. Reidel Publishing Co.
- Murphy, Jeffrie G. (1973) "Marxism and Retribution." *Philosophy & Public Affairs* 2(3), 217-43.
- Murphy, Jeffrie G. (1971) "Three Mistakes about Retributivism." *Analysis* 31, 166-69.
- Nagin, Daniel S. (1998) "Criminal Deterrence Research at the Outset of the Twenty-First Century." *Crime and Justice* 23, 1-42.
- Nozick, Robert. (1981) *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.
- Nozick, Robert. (1974) *Anarchy, State, and Utopia*. New York City: Basic Books.
- Pasternak, Avia. (forthcoming) "Sharing the Costs of Political Injustices." *Politics, Philosophy, and Economics*.
- Pasternak, Avia. (2011) "The Distributive Effect of Collective Punishment." In Tracy Isaacs and Richard Vernon, eds. *Accountability for Collective Wrongdoing*. Cambridge University Press.
- Pengelly, Gaynor. (2010) "Give prisoners bank accounts." *This is Money*. Online at http://www.thisismoney.co.uk/savings-and-banking/article.html?in_article_id=517136&in_page_id=7&position=moretopstories (accessed Oct. 25, 2010).
- Perkins, Lisa H. (1970) "Suggestion for a Justification of Punishment." *Ethics* 81(1), 55-61.
- Pettit, Philip. (2007) "Responsibility Incorporated." *Ethics* 117: 171-201.
- Pettit, Philip. (2003) "Consequentialism." In Stephen Darwall, ed. *Consequentialism*. Malden, Mass.: Blackwell Publishing.
- Pincoffs, Edmund L. (1977) "Are Questions of Desert Decidable?" In J.B. Cederblom and William L. Blizek, eds., *Justice and Punishment*. Cambridge, Mass.: Ballinger Publishing Co.
- Primoratz, Igor. (1989) "Punishment as Language." *Philosophy* 64, 187-205.

- Rawls, John. (1964) "Legal Obligation and the Duty of Fair Play." In S. Hook, ed. *Law and Philosophy: A Symposium*. New York City: New York University Press.
- Reiman, Jeffrey. (1985) "Justice, Civilization and the Death Penalty: Answering van den Haag." *Philosophy & Public Affairs* 14, 115-48.
- Rhodes, Lorna. (2004) *Total Confinement: Madness and Reason in the Maximum Security Prison*. Berkeley, Calif.: University of California Press.
- Ross, W. D. (1930) *The Right and the Good*. Oxford: Oxford University Press.
- Sadurski, W. (1985) *Giving Desert its Due*. Dordrecht: Reidel.
- Scheid, Don E. (1997) "Constructing a Theory of Punishment, Desert, and the Distribution of Punishments." *Canadian Journal of Law and Jurisprudence* 10, 441-506.
- Shafer-Landau, Russ. (2003) *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shafer-Landau, Russ. (2000) "Retributivism and Desert." *Pacific Philosophical Quarterly* 81, 189-214.
- Sher, George. (1989) *Desert*. Princeton, N.J.: Princeton University Press.
- Simmons, A. John. (1996) "The Principle of Fair Play." In Robert M. Stewart, ed. *Readings in Social & Political Philosophy*. New York City: Oxford University Press.
- Sterba, James. (1977) "Retributive Justice." *Political Theory* 3, 349-362.
- Stilz, Anna. (forthcoming) "Collective Responsibility and the State." *The Journal of Political Philosophy*.
- Tawie, Sulok. (2007, July 23) "Now, case of excessive caning by teacher." *New Straits Times* (Malaysia).
- Ten, C. L. (1990) "Positive Retributivism." *Social Philosophy and Policy* 7(2), 194-208.
- Tonry, Michael. (1994) "Proportionality, Parsimony, and Interchangeability of Punishments." In R. A. Duff and David Garland, eds. *A Reader on Punishment*. Oxford: Oxford University Press.

- U.S. Department of Justice. (2000) "Correctional Populations in the United States, 1997." *Bureau of Justice Statistics*. Online at <http://bjs.ojp.usdoj.gov/content/pub/pdf/cpus9701.pdf> (accessed Sept. 13, 2010).
- Vetlesen, Arne Johan. (2008) "Collective Evildoing." In Renee Jeffery, ed. *Confronting Evil in International Relations: Ethical Responses to Problems of Moral Agency*. New York City. Palgrave Macmillan.
- Von Hirsch, Andrew. (1999) "Punishment, Penance and the State: A Reply to Duff." In Matt Matravers, ed. *Punishment and Political Theory*. Oxford, U.K.: Hart Publishing.
- Von Hirsch, Andrew. (1993) *Censure and Sanctions*. Oxford, U.K.: Clarendon Press.
- Von Hirsch, Andrew; Bottoms, Anthony E.; Burney, Elizabeth; and Wikstrom, P-O. (1999) *Criminal Deterrence and Sentence Severity: An Analysis of Recent Research*. Oxford: Hart Publishing.
- Varden, Helga. (2008) "Kant's Non-Voluntarist Conception of Political Obligations: Why Justice is Impossible in the State of Nature." *Kantian Review* 13(2), 1-45.
- Villa, Judi. (2004, Nov. 14) "Adult Prisons Harden Teens: Young offenders groomed for life of crime." *The Arizona Republic*.
- Wood, David. (1997) "Reductivism, Retributivism, and the Civil Detention of Dangerous Offenders." *Utilitas* 9(1), 131-146.
- Young, Iris Marion. (2004) "Responsibility and Global Labor Justice." *The Journal of Political Philosophy* 12: 365-388.
- Zaibert, Leo. (2006) *Punishment and Retribution*. Aldershot, U.K.: Ashgate Publishing.