

11-1-2016

## Criminals behind the Veil: Political Philosophy and Punishment

Chad Flanders

Follow this and additional works at: <https://digitalcommons.law.byu.edu/jpl>

 Part of the [Criminal Law Commons](#), [Criminology and Criminal Justice Commons](#), and the [Law Enforcement and Corrections Commons](#)

---

### Recommended Citation

Chad Flanders, *Criminals behind the Veil: Political Philosophy and Punishment*, 31 BYU J. Pub. L. 83 (2016).  
Available at: <https://digitalcommons.law.byu.edu/jpl/vol31/iss1/12>

This Article is brought to you for free and open access by BYU Law Digital Commons. It has been accepted for inclusion in Brigham Young University Journal of Public Law by an authorized editor of BYU Law Digital Commons. For more information, please contact [hunterlawlibrary@byu.edu](mailto:hunterlawlibrary@byu.edu).

# Criminals Behind the Veil: Political Philosophy and Punishment

*Chad Flanders\**

## I. INTRODUCTION

There is evidence everywhere that our criminal justice system is undergoing a crisis of practice. Increased police violence and the concomitant distrust of police in many communities, fear of aggressive enforcement tactics more generally, worries about widespread governmental surveillance and, above all, a concern with overcriminalization and mass incarceration—these are the dreary and familiar stuff of daily headlines. But, this crisis of practice in turn reflects a deeper crisis of how we theorize about criminal law. We lack, for the most part, any worked-out theory of what the policing and processing of crime should look like. Nor do we have real consensus on what things should be criminalized.<sup>1</sup> And while we have theories of punishment in seeming abundance, they address a world that is, for the most part, divorced from our current reality.<sup>2</sup>

In particular, neither theories of deterrence nor retribution can explain, in any straightforward way, what is wrong with the warehousing of millions of people in prisons and jails across America.<sup>3</sup> We live, as many have observed, in a “carceral state.”<sup>4</sup> Our theories of criminal law

---

\* Associate Professor of Law, Saint Louis University School of Law. I have written and rewritten this paper several times over the past 5 years. I am grateful to Ben Vilhauer, Danny Priel, Joshua Preiss, David Svolba, and an audience at the University of Minnesota-Mankato for comments on a much earlier, unpublished, draft. Thanks also to Steven Galoob, Josh Kleinfeld, Tommie Shelby, Erin Kelly, Vincent Chiao, and Gabe Mendlow for comments at a University of Michigan Criminal Law Theory workshop. A first try at some of these ideas can also be found in Part III of Chad Flanders, *Can Retributivism Be Saved?*, 2014 BYU L. REV. 309 (2014).

1. For the exception that proves the rule, see DOUG HUSACK, *OVERCRIMINALIZATION* (2008) (outlining the crisis of overcriminalization and proposing a theory of what things are properly criminalized).

2. I address this point in more detail in Chad Flanders & Zachary Hoskins, *Introduction*, in *THE NEW PHILOSOPHY OF CRIMINAL LAW 1* (Chad Flanders & Zachary Hoskins eds., 2016).

3. See Chad W. Flanders, *What is Wrong with Mass Incarceration?* (Sept. 5, 2015) (unpublished manuscript) (on file with author). See also Vincent Chiao, *Mass Incarceration and the Theory of Punishment*, CRIM. L. & PHIL. 1 (2015) (arguing that major theories of punishment unable to give a satisfactory explanation of what is wrong with mass incarceration).

4. See, e.g., Katherine Beckett & Naomi Murakawa, *Mapping the Carceral State*, 16

and punishment, far from giving us grounds to criticize that practice, could in some ways be said to be propping it up.<sup>5</sup> If we need to punish harshly because retribution demands it, or because deterrence requires it, then the carceral state is not an unfortunate reality. It may be a normative necessity.

There are piecemeal solutions to the problem of the carceral state that have been offered. They are important, but I do not want to present another piecemeal solution in this essay, things that simply work at the margins. Rather, what we need now is a comprehensive diagnosis of our present situation—what has gone wrong, and how to fix it. And the way to begin a comprehensive diagnosis is to turn to political philosophy and its relative neglect of problems of criminal law and criminal punishment. Political theory would seem to be the place to address questions of the power of the state and the limits of that power, especially when it comes to the use of force against its own citizens. It would seem, then, that we might turn to political philosophy to get a grip on what is wrong with living in a carceral state.

But political philosophy has largely been silent on these topics.<sup>6</sup> This silence is especially glaring in the case of that figure that may be said to be the dominant political philosopher of the twentieth century, and who will serve as the focal point of my discussion: John Rawls. Rawls set the agenda of political philosophy in the late twentieth century and still sets it now, well into the twenty-first, especially in his book *A Theory of Justice*.<sup>7</sup> Over the years, some have tried to enlist Rawls in the project of thinking critically and creatively about criminal justice and punishment. However, they have been largely unsuccessful in using Rawls to set a broader agenda for a comprehensive look at how and why we punish and how we might think of punishment differently. This paper tries to do precisely that, but we first must be open about how hard Rawls makes things for us.

It is not that Rawls says next to nothing about punishment in *A Theory of Justice*. It is that the next to nothing he *does* say is awfully depressing and as result has been mostly ignored.<sup>8</sup> But the passages

---

THEORETICAL CRIMINOLOGY 221 (2012).

5. For this worry, especially in regard to retributive theory, see my discussion in Chad W. Flanders, *Retribution and Reform*, 70 MD. L. REV. 87 (2010).

6. See Flanders & Hoskins, *supra* note 2.

7. JOHN RAWLS, *A THEORY OF JUSTICE* (1971).

8. In her otherwise impressive paper (to which I am in debt), Sharon Dolovich ignores the passages where Rawls does talk about punishment, and develops a Rawlsian theory of punishment in spite of them. This is probably a wise move (as I argue in this paper), but it does not,

where he does talk about punishment are worth paying attention to (and being depressed about) for at least two reasons. The first is that on their face these passages seem pretty harsh: the passages seem to condemn certain people (criminals) to “misfortune” based on “their *natures*” or “the kind of persons” that “they *are*.”<sup>9</sup> In one place Rawls suggests that the only solution for dealing with such people (whose nature is their misfortune) is to *ramp up* the punishments for them until those who do not manage well in the well-ordered society are pummeled into complying.<sup>10</sup>

The second reason is that the passages are not only harsh taken by themselves; they seem especially out of place given the tone of *A Theory of Justice* as a whole. “In justice as fairness,” Rawls says at one point, we “share one another’s fate.”<sup>11</sup> Such expressions of shared citizenship and solidarity do not pop up when Rawls turns to those who violate natural duties of justice. *Their* bad acts are not the product of social circumstance and accident;<sup>12</sup> instead, they reveal the “bad characters” of those who commit such acts.<sup>13</sup> Their fate we do not share; them, we punish.<sup>14</sup>

Maybe Rawls felt his theory led him to write these passages and so he could not help them. Or maybe the passages were just tossed off and not given much thought. In either case, when it comes to punishment, Rawls’s greatest “sin” may really be one of omission. He does not treat punishment at great length—just a few pages in all—and probably is at least in part responsible for the glaring lack of treatment of punishment in political philosophy<sup>15</sup> over the last several decades.

But Rawls is also guilty of *committing* some serious “sins” in these

---

I think, tell the full story. See Sharon Dolovich, *Legitimate Punishment in Liberal Democracy*, 7 BUFF. CRIM. L. REV. 307, 326 n.40 (2004) (dismissing the few passages where Rawls discusses punishment).

9. RAWLS, *supra* note 7, at 576 (emphasis added).

10. *Id.*

11. *Id.* at 102.

12. See, e.g., *id.* at 104 (“The assertion that a man deserves the superior character that enables him to make the effort to cultivate his abilities is equally problematic; for his character depends in large part upon fortunate family and social circumstances for which he can claim no credit.”).

13. *Id.* at 315.

14. See generally Bonnie Honig, *Rawls on Politics and Punishment*, 46 POL. RES. Q. 99 (1993). Honig is more insistent than I am that these passages are essential to Rawls’s overall position in *Theory*.

15. Not in philosophy *simpliciter*—there they have dozens of essays and books on philosophy of punishment over the years. But they are not part of a *systematic* theory of justice. My hope is that this is starting to change.

passages. Those passages read in context really do show a stark divide that Rawls paints between the distributive and the retributive, one which makes his passages on punishment especially off-putting and strange. When he talks about distributive justice, Rawls is open, expansive, unifying—and radical. However, when it comes to retributive justice, Rawls is exclusionary, unyielding, and harsh. He sounds very traditional, old-school. There is at the very least a tension here, and the question is whether we can reconcile the tension, or whether another strand has to give way.

That is the project of this paper: to figure out how to reconcile Rawls on punishment with Rawls on distributive justice—and to use that reconciliation as the beginnings of a broader framework within which to think about our current crisis of criminal justice. That is, the paper is not meant to be merely yet another exercise in Rawls' exegesis, although it is partly that. It is also meant to be a map that shows *how we got here* and *how we can get out*. Thus, I want to see Rawls' work to be representative of a certain blindness in political philosophy—a blindness that led political philosophers to neglect criminal justice and punishment in designing their theories. But the way out of that is to do better political philosophy—not to leave political philosophy behind.<sup>16</sup>

In this regard, I find Rawls' brief remarks on punishment telling because, taken a certain way, his remarks can be seen to fit into a familiar ideological pattern: welfare for the working poor, and prison for those who don't work or cannot work and who fall into crime.<sup>17</sup> Rawls is not responsible for this way of dividing the world and he certainly did not invent it, but his theory does seem at points to have a place alongside it. And not unrelatedly, his book has, if only tacitly, given permission for philosophers to separate distributive and retributive justice, and mostly to *not* think about the latter in political philosophy.<sup>18</sup> So at best, Rawls has given philosophers leave to separate pun-

---

16. See Chad Flanders, *In Defense of Punishment Theory, and Contra Stephen*, 10 OHIO ST. J. CRIM. L. 243 (2012).

17. As many have noted, education/welfare and prison are both *state* responses to crime. For an extended meditation on this point, see BERNARD E. HARCOURT, *THE ILLUSION OF ORDER* (2001). For a recent discussion, see Alex Lichtenstein, *Mass Incarceration Has Become the New Welfare*, ATLANTIC (Sept. 16, 2015), <http://www.theatlantic.com/politics/archive/2015/09/mass-incarceration-has-become-the-new-welfare/404422/>.

18. This state of affairs seems to me to be a historical anomaly, but I won't press that point here.

ishment from political philosophy. At worst, his theory gives us a seemingly harsh view of punishment.<sup>19</sup>

My paper has two parts. In the first part (Section II), I draw out the various threads in Rawls that lead him to his statements on punishment. In some respects, this is well-trodden ground, especially recently. First patch of well-trodden ground: Rawls is doing ideal theory. Punishment does not fit in ideal theory except in a very limited way, and when real punishment comes in, it is only at the edges of ideal theory and not as part-and-parcel of it. Second patch: Rawls thinks that distributive justice is not about desert, but retributive justice probably is, so we have a basic asymmetry which our theory has to account for and reflect. These are familiar topics in the literature. I want to lay them out, in the first part, sympathetically—maybe *too* sympathetically—before I try to show how they tend to look when set against a backdrop of what people in the wider culture were saying about punishment and welfare, and how Rawls strangely fits that mold, even when we can see him as breaking the mold in political philosophy more generally.

The second part of my paper (Section III) lays out an argument in roughly two steps. The first step puts Rawls on punishment against the backdrop of Rawls on distributive justice to show the tensions within *Theory* both in terms of the text and in terms of the tone of the text. Pitting Rawls on punishment against Rawls on distributive justice makes Rawls on punishment seem conventional and conservative where his *Theory* as a whole is anything but.<sup>20</sup> The second step asks what it would take to read and interpret Rawls on punishment in a way that would make it consistent with other parts of Rawls' theories. This turns out to make Rawls a sort of radical on punishment, in the same way he *was* radical on distributive justice.

At the end of the day, Rawls simply does not say much about punishment in *Theory*. We are left with hints, from which we must make guesses. But there are two ways we can go. We can mostly ignore what Rawls does say, and treat it as sort of off-shoot of some methodological choices Rawls made. That is certainly one thing we can do, and in Part I of my paper, I show how we can fairly disregard Rawls on punishment

---

19. For the connection between retributive theory and harsh punishment (and the varieties of ways punishment can be “harsh”), see my discussion in Flanders, *supra* note 5, at 87.

20. Alan Gewirth wrote that his political theory was an attempt at an institutionalization of love. I don't think this is too far off of what Rawls was doing in *Theory*. See RAWLS, *supra* note 7, at 105 (Rawls on fraternity).

and still feel OK about reading *Theory* as the great work on distributive justice it is. The other thing we can do is to try to figure out a theory of punishment that fits better with *Theory* as a whole, one that tries to imagine applying the open and expansive theory of distributive justice even to those who commit crimes. I find this latter path a much more promising—and exciting—way to go. It gives us hope that we can think about our criminal justice system not just in small ways, but in big ways: ways that point out how we can reform our institutions for the better.

## II. LIMITING THE DAMAGE OF RAWLS ON PUNISHMENT

My aim in this part is to look at Rawls's statements on punishment as things that are not essential to his theory, and even as positively extrinsic to his project in *Theory*. Rawls warns us, early on, that he cannot talk about punishment if his project is one of "ideal theory." Punishment does emerge in *Theory*, but only on the margins. We make a mistake—I suggest on this sympathetic reading of Rawls—if we read his by-the-way comments on punishment as somehow disclosing deeper truths about his theory (I take this back a little in the next part of my paper). If the reading in this part is defensible, and I think it is for the most part, then excising most of what Rawls says on punishment would still leave the structure and most of the content of justice as fairness intact. We might lament that Rawls then has really *nothing* to say about state punishment in *Theory*; but this may be better than interpreting him as saying wrong things about punishment in *Theory*.

I am going to argue that, read generously, Rawls's remarks on punishment are really not about implementing punishment, or how punishment would look in a non-ideal world. Rather, what he says on punishment is introduced at certain places to simply make a point about some other part of his theory or to make a necessary contrast. They are not positive statements about punishment at all.

### *A. Ideal Theory and Punishment*

Rawls's very first mention of punishment in *Theory of Justice* is, interestingly, in a list of topics that he would *not* be discussing. The subject of his book, Rawls says, is what a perfectly just society would look like. And if this is his subject, then there will be of necessity certain areas that he will not be covering because they involve issues that

arise only when people are acting unjustly to one another. If we are dealing with a society where everyone complies with the law, then we will not be dealing with various problems when only some people comply with the law. These are topics such as “the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and militant resistance to revolution and rebellion.”<sup>21</sup> Rawls goes on to include in this category of things not part of his subject “questions of compensatory justice and weighing one form of institutional injustice against one another.”<sup>22</sup> Rawls does not deny these subjects are important; they are just not what he is dealing with in *Theory*. He wants a more systematic grasp of justice, and this means assuming full compliance.

Another name for what Rawls says he is doing is “ideal” theory, and issues of less than full compliance would then fall into “non-ideal theory.” We first get the idea by assuming what the principles of justice would be if everyone followed the principles of justice; we then go on to treat other, non-ideal situations given the ideal principles of justice. The ideal/non-ideal distinction has been the subject of much (confusing) debate,<sup>23</sup> but here Rawls’s point seems simple enough. If we are going to be talking about what a perfectly just society would look like and we assume that such a society means everyone complies with the principles of justice, it follows that we won’t be dealing with what happens when people *don’t* follow the law.

Taken as simply a stipulation, it is hard to quibble with this: if Rawls wants to leave some topic (or group of topics) for another day, then he can. We can say that this is the wrong starting point, or that it does not give us good principles of justice. But simply as a way of organizing his thoughts, it is not something that Rawls can really get right or wrong. It is what Rawls wants to do. One might also criticize Rawls for developing principles that no person could follow; but this is a criticism in a different register. *That* would be a criticism that Rawls was not even doing ideal theory well. But it is not a criticism just to say that Rawls is not saying what happens or how we are to decide

---

21. *Id.* at 8.

22. *Id.*

23. For an overview, see Laura Valentini, *Ideal vs. Non-ideal Theory: A Conceptual Map*, 7 PHIL. COMPASS 654 (2012). See also Chad Flanders, *The Mutability of Public Reason*, 25 RATIO JURIS. 180 (2012) (presenting thoughts on ideal theory in a slightly different context).

things when people break the rules—again, he is not *trying* to say anything about those things.<sup>24</sup>

Rawls returns later to the topic of punishment in *Theory*, so there is some sense in which his initial take at separating punishment from strict compliance theory fails. He said he could not talk about punishment because he was doing ideal theory, and then it turns out he does talk about punishment, *and even at times as part of an ideal theory*. But it is important to see, at least at first, and if we are trying our best to be charitable, how Rawls's subsequent discussions of punishment do not mean he is going against his early promise to stick with a theory of justice with the assumption of strict compliance. For it turns out that even though the ideal society *will* have a system of punitive sanctions, the sanctions in an “ideal” society will play a different role than they would in a society that was marked by law-breaking. And though it is tempting to read all the passages where Rawls talks about punishment as Rawls's attempt to lay out a non-ideal theory, we should resist this temptation, and heed Rawls's original admonition that he is doing only ideal theory.

This point is clear in Rawls's first extended discussion of punishment in *Theory*, where he is explicit about the role of penal sanctions even in ideal theory.<sup>25</sup> Here is how Rawls's argument goes. When we enter society, we do not know one another's motives.<sup>26</sup> While we may know ourselves, or think we do, and know that we are motivated by our sense of justice to obey the law and follow the principles of justice, we do not know that this is the case with others.<sup>27</sup> They may not have a fully developed sense of justice, and so may just be looking for the first chance to break the law and take advantages of the compliance of others. Or perhaps we may even suspect that we ourselves at times may succumb to temptation. If our obligation to obey the law was purely voluntary, and not a matter at least to some degree of not risking punishment, we may fear that the whole system would fall apart.

In short, we might like there to be sanctions, either to boost our own possibly flagging motivations, or to make us feel better that others will be motivated to obey the law. If it were just a matter of following

---

24. At the same time, my essay does end up making a version of this criticism: Rawls leaves out punishment as part of his political theory, and this is a very serious failing. But at the same time, I see it more as an expression of disappointment, not a criticism *per se*.

25. See RAWLS, *supra* note 7, at 240–42.

26. See *id.* at 240.

27. *Id.*

the law out of *preference*, we might think others will all too easily break the law when it suits them. And we might worry that we would be tempted to do the same thing.<sup>28</sup>

If this is accurate, we might need to have a system of punishment set up for the sake of our sense of security: that others will really live up to their obligation to obey the law—and that we will too. We need the assurance, that if people’s sense of justice fails, then at least there is a “back-up” system of incentives to (in the Kantian jargon) act according to duty even if not to act from duty. And we need this system even if we postulate that everyone is going to obey the law and so the penal sanctions will never need to be used.

That is to say, we need the system of penal sanctions even in a fully just society, one marked by full compliance, so even “ideal theory requires an account of penal sanctions as a stabilizing device . . . .”<sup>29</sup> It is, again, a matter of ideal theory because it shows how a system of punishment will exist even in a society where by hypothesis everyone obeys the law. It is not about what to do when someone breaks the law. *Ex hypothesi*, in an ideal society no one breaks the law and no one is actually punished. Yet we still have a punishment system, the existence of which “serves as men’s security to one another.”<sup>30</sup>

Keeping this in mind can also help us interpret Rawls’s later treatment of punishment and ideal theory, which is where some of his harsher remarks appear. In Section III of *Theory*, Rawls raises the important question: what is the relationship of our good, our happiness, to justice?<sup>31</sup> What if it turns out the principles of justice ask us to live in a way that, for most of us, makes us miserable?<sup>32</sup> Rawls wants to assure us that this will not be the case and that the “strains of commitment” of justice as fairness are less, if not non-existent, especially as compared to what utilitarianism may ask of us.<sup>33</sup> The principles of jus-

---

28. In what follows, I am indebted to Paul Weithman’s excellent discussion of stability and sanctions in PAUL WEITHMAN, WHY POLITICAL LIBERALISM?: ON JOHN RAWLS’S POLITICAL TURN 50–51 (2010). Weithman is discussing Hobbes, but Rawls indicates his debt to Hobbes in this part of RAWLS, *supra* note 7, at 240 n.24.

29. RAWLS, *supra* note 7, at 241.

30. *Id.* at 240.

31. *Id.* at 395. The discussion of “the good of justice” for a person takes place in Chapter IX of *Theory*.

32. “The problem,” Rawls says in one passage, is whether the desire to “adopt the standpoint of justice belongs to a person’s own good.” *Id.* at 568.

33. *Id.* at 572–73.

tice, Rawls wants to argue, are both individually and collectively rational.<sup>34</sup> They are good for society—the form of society they present is a good one—and they are good for the individual person.

But in an aside, Rawls asks what if, “even in a well-ordered society there are some persons for whom the affirmation of their sense of justice is not a good”?<sup>35</sup> In the case of a person who does not see following his sense of justice as a good for him, there is a lack of congruence. He *will* find obeying the law as a real burden because he is not realizing his nature. Of him, Rawls says that his “nature is his misfortune.”<sup>36</sup> It is his misfortune, because the principles of justice still remain justified, and so in the abstract worth following and collectively rational for all that, even if they are not good for *him*. They are the principles of justice we would all choose behind the veil of ignorance, and it is his bad luck that upon removing the veil, he chafes against the principles of justice rather than finding them at least consistent (if not congruent) with his good.

We cannot, Rawls says, let this person hijack the collectively rational principles of justice and tailor them towards *his* good, to allow him to break the law because he is unhappy.<sup>37</sup> But, we have to be careful here. Rawls is not saying that there *would* be people like this in the well-ordered society. All he is saying is in the conditional. *If* there is a person like this, the principles of justice—it almost follows—would be bad for him because we are dealing with precisely of the (hypothetical) case of a lack of congruence between a person’s good and the principles of justice.

Still in the conditional: What if a *lot* of people in the well-ordered society are like this? Now, Rawls cannot rule this out altogether; presumably his hope is that the well-ordered society will produce people in which the sense of justice is consistent with their good, but maybe this is not a reasonable hypothesis about the good of human sociability.<sup>38</sup> But even in the case where justice is not a good for most people,

---

34. *Id.* at 577.

35. *Id.* at 575.

36. *Id.* at 576.

37. *Id.* at 575–76 (rejecting the idea that a person can be required to do “only what is to his advantage”).

38. *See id.* at 522–23, 527 (expositing the good of human sociability). It is possible that Rawls thought that a well-ordered society would “solve” crime. After all, the people in the well-ordered society will have been *raised* in a well-ordered society, and so had ample opportunity to cultivate their sense of justice. They would also, presumably, have “enough” primary goods so they would not steal or commit other crimes in order to live decent lives. I do not think Rawls held this position, although I don’t pursue the point at any length here.

Rawls will still say that the principles of justice are *still* collectively rational. And so in *that* society—where there are a lot of people who find following the principles of justice makes them miserable—we will have to rely on penal sanctions a lot more. Again, we need to be careful here. Rawls is not talking about actually using penal sanctions on anyone. In fact, Rawls does not really say much about actually employing penal sanctions *and he should not* because his project is a project in ideal theory. But even an ideal society will have to rely at least on the threat of penal sanctions as a stabilizing mechanism. We need these mechanisms not because we want to punish people, but we want to have an extra assurance that people, including ourselves, will obey the principles of justice.

Summing up, “penal devices” in a society with a lot of people who do not find their good in being just will play a much larger role not necessarily because a lot of people will be punished, but because a lot more people will find their motivation for acting according to the principles of justice underwritten by the fear of sanctions rather than by a love of justice. The fear of punishment, not actual punishment, will loom larger in the minds of many in the society. However, the purpose of the sanctions is not in the first instance to punish, but rather to “underwrite citizens’ trust in one another.”<sup>39</sup> It is about how we relate to one another as citizens in a well-ordered society, not how some people punish other people when they break the law.

We are still in the land of ideal theory. We are not talking about what happens to people who break the law, but about what to do with people who may be less than perfectly motivated to obey the law. A society in which “penal sanctions play a much larger role” *could still be a system in which all people comply with the law*. And it had better be, for it to still be a wholly “ideal” society. It is just that for some people in that society, reasons for compliance will not be their sense of justice, but their fear of punishment. Such a society will probably be a lot less stable than one in which people mostly act out of their love of justice.<sup>40</sup> There will probably be more mistrust, and more need to point to the existence of a system of penal sanctions. At the limit, if the *only* reason *anyone* was behaving justly was because of the fear of punishment, we might have to rethink the theory: the strains of commitment would be too great, and the whole system could collapse. But

---

39. *Id.* at 575.

40. See WEITHMAN, *supra* note 28, at 50–51.

Rawls does not think this is a real possibility although, I suppose, he cannot rule it out in principle.

What is interesting on this interpretation of Rawls's remarks is that we are only talking about the threat of punishment. Nothing in what he says explicitly deals with what we do with people who fail to obey that threat. It is possible that people do not get punished, but treated. The ideal society needs punishment, but only as a sort of signal to the members of society that if their fellow citizens aren't moved by justice there are other things that are there to move them. Whether or not we have to make good on what those threats say, as H. L. A. Hart showed in his discussion of Bentham, is another question.<sup>41</sup> That we have a system of threats of sanctions does not mean that we have a system of actually employed sanctions. Whether and how we impose those sanctions is, in the end, a question for non-ideal theory.

### *B. Distributive Versus Retributive Justice*

The distinction Rawls makes between distributive and retributive justice is a little more familiar, and so needs less exposition.<sup>42</sup> My point here is not necessarily to say anything new, but to see the contrasts Rawls draws (between ideal and non-ideal, between retributive and distributive justice) as not really a part of his main theory, and as putting that theory into a sort of relief. Again, I am trying to read Rawls charitably in this part. I do not want to identify him with the harsh rhetoric and its implications if I can help it—at least initially. Above, I tried to do this by saying that the language of ramping up sanctions and of some people having the “misfortune” to have a certain nature as applying only to the existence of sanctions to make sure people will follow the law. It did not mean that there would be any such people, or even that sanctions would be used. It is harder to do with Rawls's remarks on retributive justice, where he seems to endorse that as the purpose of punishment, but it is possible.

Rawls introduces the contrast between distributive and retributive

---

41. Hart's argument strictly speaking runs the other way around: that the fact that a threat is ineffectual does not mean that there is not any value in imposing the punishment. See H. L. A. HART, PUNISHMENT AND RESPONSIBILITY 18–20 (1968). The larger point, I take it, is that the threat and the punishment can be analytically separated. Many have tried to justify the imposition of punishment based on the idea of legitimate threats, but it has always been a problem with these theories to transition from the justified *threat* of punishment to the justified *imposition* of punishment. See, e.g., WARREN QUINN, MORALITY AND ACTION 52 (1993).

42. See Chad Flanders, *Can Retributivism Be Saved?*, 2014 BYU L. Rev. 309 (2014).

justice to resist the idea that, like retributive justice, distributive justice should be based on merit. But Rawls says that this contrast is fundamentally misguided. There really is no such thing as a pre-institutional *economic* desert. In fact, the very notion is fantastical. Economic desert, if it exists at all, is the upshot of a “scheme of taxes and burdens designed to put a price on certain forms of conduct and in this way to guide men’s conduct for mutual advantage.”<sup>43</sup> In the economic realm—the scheme of taxes and burdens—what we have are legitimate expectations created by institutions, not anything that we can be said to pre-institutionally deserve. The economic sphere creates the expectations; it does not exist to satisfy pre-existing claims of desert.

Contrast this, Rawls says, with how we think of the criminal law. There, we *do* have a sense of right and wrong, and accordingly of desert. Those who have violated the criminal law have normally, Rawls says, done something that we can see as *wrong* apart from the existence of natural duties.<sup>44</sup> They have injured someone or killed someone. Punishments regulate something that it would be if they “were never done.”<sup>45</sup> Moreover, those who do the bad things—that the criminal law prohibits—really do show “a mark of bad character.”<sup>46</sup> If they show these characters by doing bad acts, they really do deserve to be punished.<sup>47</sup> It would be perverse, Rawls seems to be saying, if we saw punishment as just a system of “taxes and burdens” to regulate conduct.<sup>48</sup> No, there are some things that are simply wrong to do and those things deserve to be punished. There is a fundamental contrast between the distributive and the retributive.

It is tempting—and mostly possible—to read what Rawls says here as not *really* in his voice. It is tempting, that is, to read it in the following way: “A lot of people may see retributive justice as pointing to the way in which we should also think of distributive shares. People see the criminal law and punishment as serving a function that is based on desert. *Whether or not that is the case*, the contrast is a mistaken one.” In other words, Rawls can be read not as endorsing a theory of retributive justice, where people deserve to be punished, but as simply putting such a theory in the mouth of an opponent. His broader point is

---

43. RAWLS, *supra* note 7, at 314–15.

44. *Id.*

45. *Id.* at 315.

46. *Id.*

47. *Id.*

48. *Id.* at 314–15.

that we should not think of economic desert as the basis for a system of distributive justice. He does not need to believe in the theory of retributive justice he lays out, only that the “opinion” of the person who thinks that because retributive justice is one way (punishment according to desert), distributive justice has to be that way, too (economic rewards according to desert).

We can go a lot of the way toward ascribing this sort of strategy to Rawls, although maybe we cannot go *all* the way.<sup>49</sup> It is hard to say with any real certainty what Rawls is doing here. Rawls does seem at points to be speaking in his own voice. One place where this seems especially clear is in how he links up punishment to his theory of natural duties of justice. Rawls says, “the purpose of the criminal law is to uphold basic natural duties . . . .”<sup>50</sup> These *are* duties that exist prior to any institutional scheme. We have an obligation to do them, regardless of whether we are part of the “scheme of taxes and burdens”<sup>51</sup> or not. And Rawls is correct that these are things that it is always the case that it would be “far better if . . . [they] were never done.”<sup>52</sup> But it is unclear that believing this entails anything like saying that those acts are marks of a bad character, which is what Rawls goes on to say. (As we will see in the next part, saying this seems to be inconsistent with what Rawls says elsewhere in *Theory* about what goes into a person’s character.) In the next paragraph, Rawls reverts again to the ideal/non-ideal distinction when he says that punishment in ideal theory is about assurance, not about actual punishment.<sup>53</sup> For the most part, Rawls goes on to say, criminal justice is for partial compliance theory, and it does not follow from this that we have to believe anything about character or desert in punishment given what Rawls says about punishment in ideal theory.<sup>54</sup>

And such a reading of Rawls—where he is not committing himself to a theory of punishment based on desert—fits better with what Rawls had said earlier about punishment. In his first discussion of the ideal theory of punishment, where he says that even ideal theory needs a limited account of punishment, Rawls says that looking at the limited way punishment operates in ideal theory we can “see that the principle

---

49. In fact, it is in these passages that my criticism of Rawls gets its foothold.

50. *Id.* at 314.

51. *Id.*

52. *Id.* at 315.

53. *Id.*

54. *Id.*

of responsibility is not founded on the idea that punishment is primarily retributive or denunciatory.”<sup>55</sup> Rejecting punishment as retributive or denunciatory is consistent with saying that punishment is still justified, but it seems incompatible with saying that punishment is a matter giving “bad characters” what they deserve. We need punishment, as a threat and maybe as eventually an applied sanction, to assure one another that there are incentives for compliance, incentives that can fill whatever gap there is between peoples’ sense of justice and their actual behavior. Rawls even in these pages says that it is as a *stabilizing device*—and not, say, a device for securing “just deserts”—that the idea of punishment first gets its foothold in his theory.<sup>56</sup> We simply need the rules to guarantee compliance, not to serve any further end, like giving bad characters their due. Thus, we might read Rawls’s remarks on retributive justice as mostly setting up a contrast with distributive justice, and perhaps hinting at a non-ideal set-up—but *not* as presenting his own “positive” retributive view on punishment.

### III. WHAT’S STILL WRONG WITH RAWLS ON PUNISHMENT— AND HOW TO FIX IT

I have tried, in the previous part, to read Rawls in a way that gets him off the hook for some of his harsher remarks on punishment—about whom we punish and why we punish them. We can read his statements as not really in his own voice, or at worst speculations about what a non-ideal theory might look like that we can take or leave. Generally speaking, even if they are in his own voice, there is not much here to go on, so we should be wary of trying to divine a fully-worked-out theory of punishment somehow implicit in what he says mostly in passing. There is no theory there. That is bad enough, but maybe not *as* bad as a worked-out theory that is harsh or essentialist or otherwise objectionable.

In this part, I still want to say that there are problems with the way punishment is treated by Rawls in *Theory*. Maybe the biggest problem is the fact that the book has so little to say about punishment. I think this silence says a lot, and it also has probably set the agenda for much of political theory: for most of the last fifty years, political theorists have been debating distributive justice and not the state’s role in pun-

---

55. *Id.* at 241.

56. *Id.*

ishment, which is certainly the most obvious and most worrisome instance of state coercive power.<sup>57</sup> And Rawls continued *not* to talk about punishment in *Political Liberalism*, even though that work is in many ways quintessentially non-ideal: it responds to real world conditions (in particular, the existence of “reasonable pluralism”) and how they limit the possibilities of overlapping consensus.<sup>58</sup> *Political Liberalism* is also historical in a way that *Theory* is sometimes not. I suspect (and have not been the only one to suspect) there is a sort of ideological blindness to the very way *Theory* is structured, in that it deals with ideal theory rather than non-ideal theory.<sup>59</sup>

*At the very least*, it is a missed opportunity. Rawls sets up a compelling and inspiring way of looking at society in *Theory*. It is a model of “fraternal” cooperation, where we are linked to our fellow citizens in a “social union of social unions.”<sup>60</sup> When we leave out those we punish—both in the set up of the theory and in its “ideal” application—we implicitly say that those who break the law cannot be part of this social union. They are banished, as it were. This is not to say that there are problems—theoretical and practical—with the idea that criminals are also fellow-citizens.<sup>61</sup> But Rawls as a matter of the structure of his theory puts this problem to one side and thus in a way answers it. So in this part I do two things: (1) I show how different Rawls is on distributive justice, as opposed to the way he discusses criminal justice (when he does discuss it), and (2) I try to imagine how a Rawlsian theory of punishment would look, if we ignored the way Rawls brackets problems of criminal justice. I will admit I am not the only person who has tried the latter; more than a few people have.<sup>62</sup> But for one reason or the other, the idea has not stuck.

---

57. Again, my point is not that there has not been philosophy of punishment, but that there has not been much *political philosophy of punishment*.

58. JOHN RAWLS, *POLITICAL LIBERALISM* (1993).

59. See Charles W. Mills, “*Ideal Theory*” as *Ideology*, 20 *HYPATIA* 165 (2005).

60. RAWLS, *supra* note 7, at 527.

61. For instructive thoughts on this issue, see R. A. DUFF, *PUNISHMENT, COMMUNICATION, AND COMMUNITY* (2001).

62. In addition to Dolovich, see GLEN LOURY, *RACE, INCARCERATION AND AMERICAN VALUES* (2008); PAUL BUTLER, *LET’S GET FREE* (2010). See also Samuel J.M. Donnelly, *The Goals of Criminal Punishment: A Rawlsian Theory (Ultimately Grounded in Multiple Views Concerned With Human Dignity)*, 41 *SYRACUSE L. REV.* 741 (1990).

A. *The Invidious Contrast Between Distributive and Retributive Justice*

Start with two big ideas of Rawls's theory of justice as fairness: first, the idea that for true equality of opportunity, we have to start with the assumption that who we are, our talents, are not really ours, in the sense that we really do not deserve them; and second, the idea that we should choose from behind the veil with the idea that we could be one of the people who is "the worst off." These are of course crude summaries of ideas that are expounded upon at length in *Theory*; I do not mean to have captured all of their subtlety in my paraphrase. But surely some of the power of these ideas *is* in their broad, vague appeal, and in the orientation they suggest. In a just society, we take into account that people have different starting points, and that we should not give ourselves airs that we hit a triple when we were really born on third base. Nor should we design policy just to help ourselves; we should think first about those least well off. All of this suggests a radical empathy, which I believe suffuses a *Theory of Justice* and explains a lot of its enduring appeal.

But as they are applied in *Theory*, these two ideas also show the severe limits of that empathy. Start with the idea that underlies much of Rawls's approach to desert. We are not responsible for the families we are born in, or the city, or the nation. Where we start out is a matter of luck. But Rawls takes this point and starts even further back. If we are not responsible for *where* we start out, neither are we really responsible for what we start out *with*.<sup>63</sup> We all are born with certain innate abilities and talents, which, to be sure social circumstances can help us realize or not, but which still are to some extent independent of them. This may be especially obvious in the case of certain mental and physical disabilities; these will limit people from the very beginning, and no amount of subsequently added social capital may be able to fully make up for that lack.

So Rawls designs much of his theory around the idea that because we are not responsible for our virtues, or the lack thereof, we cannot run society around any firm idea of merit or entitlement based on our

---

63. RAWLS, *supra* note 7, at 311–12.

talents. Some people may be lazy, or impatient, or worse, but this is no fault of their own. We cannot blame them for it. Neither can we credit people for their talents; these, too, are mostly a matter of luck. We should look at others and say, “There but for the grace of God (or circumstances) go I.” We are in a sense only vehicles for those talents, not deserving what results from them “any more than one deserves one’s initial starting place in society.”<sup>64</sup>

I do not want to get into whether this is an adequate theory of who we are, that is, whether it gets our identities as persons with traits right.<sup>65</sup> Nor, for that matter, am I that concerned with how this conception of our talents fits with another of Rawls’s big ideas, which is that in the economic sphere it is a bit nonsensical to say that there is really any meaningful sense of “desert.”<sup>66</sup> Certainly the idea that we don’t deserve our talents relates in *some way* to the idea that it is incoherent to talk about economic desert.<sup>67</sup> All I want to insist is that there is also a sense in which it would be wrong to give people more money just by virtue of the talents they have as a matter of sheer luck.

The point I want to make here is to note that Rawls here seems only concerned with things like laziness or impatience or general lack of talent: those kinds of less-than-optimal character traits that may make us unable to compete in the marketplace. He is *not* concerned with people who may have the kinds of character that will incline them not just to not succeed, but who may have a tendency to break the law of any given scheme of social cooperation. Of course, it is possible and even likely that those characteristics that make people not do well in the market may also incline them to criminal behavior: they may tend to discount the future consequences of their immediate actions, for example.<sup>68</sup> Rawls is not interested in those who cross the line from legal but inefficient behavior to illegal behavior.

Part of this, of course, is attributable to Rawls’s interest in ideal theory. But it is hard to avoid this implication: people who are lazy are not to be blamed for that character trait, but those who break the law are responsible for their bad characters. That is, it is not that hard to

---

64. RAWLS, *supra* note 7, at 104.

65. See, e.g., MICHAEL J. SANDEL, LIBERALISM AND THE LIMITS OF JUSTICE (1982).

66. See, e.g., SAMUEL SCHEFFLER, BOUNDARIES AND ALLEGIANCES: PROBLEMS OF JUSTICE AND RESPONSIBILITY IN LIBERAL THOUGHT (2001) (especially pressing this interpretation).

67. I go into this more in Flanders, *supra* note 42.

68. See Dolovich, *supra* note 8, at 319 for a similar point.

read Rawls's points about our natural duties and about retributive justice in this light. When we get to criminal justice, we are dealing with acts of people that really reveal their characters in what they do and those actions are not just the upshot of luck, or fate. People who break the law have a choice whether to break the law or not; not so people who lack certain talents or who are just lazy or impatient or lack the talent to compete in the marketplace. The latter group is not responsible for what they have become, and so we must share their fates.<sup>69</sup> People who break the law are responsible for what they do, and so they have only themselves to blame.<sup>70</sup>

In this way, then, those who tend to break the rules—the criminals—are left out at the front end. The accidents and social circumstances that led them to be what they are is not the fault of society, it is their fault alone, their nature is revealed in their bad acts, and their nature is their misfortune. The pieces *do* seem to fit together, if we work at it.

But there is also a way in which criminals are left out at the back end as well—and here I want to get to another of Rawls's big ideas, namely, that society should be ordered in such a way as to help out those who are the “least advantaged members of society.”<sup>71</sup> In considering those least well off, in imagining who we might be when the veil is lifted, we are not to imagine that we might be disposed to rebel against our sense of justice, or not have a sense of justice at all. *Systematically*, the original position excludes those from consideration who might be tempted to *not* follow the principles of justice when the veil is lifted. *Systematically*, we are not to direct our empathy toward them in deciding whom to designate as the “least advantaged.”

Again, this move of exclusion can be justified by Rawls's saying that he is interested in only ideal theory. He does not want people to be

---

69. See generally RAWLS, *supra* note 7, at 311–12 (“[T]he initial endowment[s] . . . in early life are arbitrary from a moral point of view. The precept which seems intuitively to come closest to rewarding moral desert is that of distribution according to effort, or perhaps better, conscientious effort. *Once again, however, it seems clear that the effort a person is willing to make is influenced by his natural abilities and skills and the alternatives open to him.*”) (emphasis added).

70. We might think this is a Kantian version of freedom: you can always choose not to break the law, so this is not a morally salient characteristic behind the veil. But why couldn't we say of the person who is lazy: you should work harder! In fact, Kant probably thought that we *could* demand this of him. But Rawls does not. See RAWLS, *supra* note 7, at 73–74. In either case, to deny that there are some social and biological determinants of criminal behavior seems to me incorrect—one cannot just “choose” not to have the sort of temperament that may lead one to break the law rather than comply with it.

71. *Id.* at 15.

picking principles of justice for those who would break the law, or designing “just” institutions around them.<sup>72</sup> As he says, it does not have to be the case that the principles of justice will answer to the good of everyone, nor is it a fair expectation that we should adjust the principles of justice to the good of those who would harm the good of other people. So we adopt the assumption that people are choosing principles for those who are going to obey them, who care about the good of others, and not just their own. We do make concessions to those who struggle in the market—this is why we have the difference principle, and not some other principle—but not those who break the law.<sup>73</sup>

Fair enough. But note how this sets up the following picture of society. Those who are within the distributive scheme, who play by the rules, even though they may struggle at times—they are part of the cooperative venture for mutual advantage. They are the people we want to see succeed, or at least manage, in society. Then there are those who are outside of it: not considered within the original position, and, when the veil is removed, not part of the cooperative venture for mutual advantage. They are not treated as needing incentives or distributive shares or anything; they get punished when they break the law because their nature leads them to break the law rather than to work within it. They do not get the benefits of those who are designated the least well off in society, even though we might think that it is an accident of nature, of fortune, that they are the way that they are. In general, justice as fairness tries to counteract the effects of chance and nature and the arbitrary hand of fate. Not here.

Might this picture seem all too familiar? Imagine a politician who says something along these lines:

There are some people who are poor, but they are the good poor, the working poor. For them, we extend to them welfare, or maybe even a guaranteed income; certainly we give them a break on the taxes they pay. They need a hand up. But then there are those who are not the good poor. They are the ones who try to get ahead by breaking the rules, by stealing, killing, or some other crime. Or they are impulsive, and their impulsiveness leads them to violate the law. It is important to punish the bad poor, those who break the law, because after all they had a choice (look at the good poor!) to do or not do what they

---

72. *But see* RAWLS, *supra* note 7, at 31.

73. As an aside, the fact that we do not think of criminals *pre-veil* does not mean that we cannot think about them *post-veil*. In fact, Rawls’s remarks on “tolerating the intolerant” shows how this can be done. *See id.* at 220.

did. And when they are done being punished, it is up to them to figure out how to get ahead in society. Moreover, they in some sense (even after they have been punished) do not deserve help from society. So, it is OK if we attach to them post-punishment disabilities that make it harder for them to get a job or to get public housing or to get education. After all, it is not our fault that they committed the crime; it is not our fault that they have the nature they do.

Again, in Part I of this essay, I tried to explain how we do not have to attribute this picture to Rawls. But nonetheless, we can read those passages in *Theory*, the ones that are so distressing, as in some way fitting this picture. There is an asymmetry between the distributive and the retributive. So long as one works hard and plays by the rules, that individual has a place in society, even if that person is not the most talented, even if they need a helping hand. But when the individual breaks the rules, our attitude towards the individual shifts, or seems to; we blame them, we hold them responsible, and the appropriate response is *punishment*, not help, not welfare. Committing crimes is a “mark of bad character” and legal punishment should fall upon the individual when they commit a crime. Insofar as following just rules is not part of that individual’s good, then, sorry, their nature is their own misfortune. Of course, I am reading into the silences here. But is the reading that far off?<sup>74</sup>

### *B. Reconstructing Rawls*

Rawls excludes criminals from his picture, those who break the law, either because for methodological reasons he excludes them, or for reasons having to do with what he thinks (but does not elaborate on in

---

74. Is it? Consider this passage from Scheffler’s essay on Rawls and distributive and retributive justice:

The problem of retributive justice is not the problem of how to allocate a limited supply of benefits among equally worthy citizens but rather than problem of how society can ever be justified in imposing the special burden of punishment on a particular human being. To put it another way, the establishment of penal institutions is a social response, not to allocative concerns, but rather to exercises of individual agency that society deems intolerable.

SAMUEL SCHEFFLER, *Justice and Desert in Liberal Theory*, in BOUNDARIES AND ALLEGIANCES 173, 192–93 (2001). Note how punishment in this picture becomes, fundamentally, about imposing sanctions on an individual human being’s exercise of his “particular” agency. But why frame things this way? Why not treat crime as a social problem that deserves a social response but not necessarily a punitive/retributive one? Asking the question *this way* I take to be a major theme of Vincent Chiao’s recent work. See Vincent Chiao, *Two Conceptions of the Criminal Law*, in THE NEW PHILOSOPHY OF CRIMINAL LAW 19 (Chad Flanders & Zach Hoskins eds., 2016).

*Theory*) about retributive justice and about people's character when they commit crimes. The sympathetic reading of Rawls attributes this exclusion to the fact that Rawls is doing ideal theory, and that sometimes he introduces retributive justice as a point of contrast—not that he believes in it. The less sympathetic reading, even the aggressively hostile reading of Rawls, has him putting criminals pretty much outside of the social contract. But whether the exclusion is methodological or substantive, there is something that the two readings share. They both say we treat criminals separately because they are in some sense *outside* of a theory of justice. They are instead to be treated as part of a non-ideal theory of punishment. The only real question that is left is whether we are on our own to divine what this non-ideal theory would look like, or whether Rawls has already dropped some hints in *Theory*.

But what if we started not with exclusion, but with inclusion? We would, for starters, remove the artificial distinction between responsibility for good and for bad character.<sup>75</sup> Put people who cannot compete in the marketplace, or who have strange or out of the ordinary conceptions of the good, on par with those who for reasons of social accident lack a sense of justice, or find their good not in following their sense of justice. Treat *all* people as not deserving their characters, all the way down, and figure out how to organize society on those grounds. And get rid of the idea that the economic sphere is a scheme where we all contribute and all agree to share one another's fates, but that the sphere of law and order is a separate one, where each one of us is individually responsible for our flaws. Finally, look at those who commit crimes, and who may suffer punishment for them, as possible objects of empathy, just as those who lack skills or talents to keep up with the Joneses. Do not look at lack of compliance as something to be treated only by ramping up sanctions to deter bad behavior. Look at the sanctions as possibly befalling fellow citizens, and draw up societal rules for law-breaking and for punishments accordingly.

If we do all this, then we can start looking at responsibility for criminal behavior as falling on all of us—or on none of us, individually—in the same way that we start looking at the economic system as all of ours, and not each of ours, individually. It is a system of incentives and prohibitions, and we try to make it work out to the benefit of all,

---

75. Michael J. Sandel goes very far in this direction, and is rather convincing (in my view). But then he flinches, and says that considerations of reflective equilibrium might come in to reintroduce the distinction between distributive and retributive justice. See MICHAEL J. SANDEL, LIBERALISM AND THE LIMITS OF JUSTICE 89–92 (1982).

especially the least well off. It is certainly true that it is better (as Rawls says) that some things, such as violations of natural duties, never happen.<sup>76</sup> But there are two ways we can look at that. We can look at it as a matter of punishing those that fail to live up to their natural duties. Or we can look at it as a collective failure, which we try to prevent, and then when it happens, we try to remedy and correct.

This latter view is what I am pushing here, because it seems to me more consistent with what *Theory* as a whole says about desert. That is, if we kick desert out of both distributive and retributive justice, then both of them become importantly demoralized. We do not reward virtue with happiness, nor do we punish bad character with punishment. Instead, we treat them both as things to be regulated as part of a larger system. We try our best to fairly distribute both happiness and punishment in ways that maximize the good of the whole, and especially maximize the prospects of the least well off.<sup>77</sup>

Of course one part of this is to set up clear rules so people know what side of the line to stay on (notice is still important). We try, that is, to avoid violations of natural duties in the first place by making clear what people should and should not do. Again, it is better that a violation of the natural duties not happen at all. And it may even be to the best that we *threaten* harsh sanctions if people fail to live up to their natural duties. But it is not obvious that what follows from violations of natural duties is *punishment*. For one, it is probably the case that those who lack a sense of justice should be counted among the least well-off. They are not able to enjoy the good of justice, which is a very great good, both individually and collectively. Their nature is, in a real sense, their misfortune on those grounds alone (this is the more optimistic way to read that passage in Rawls). Moreover, it is not obvious that punishment is the best way to maximize the good of the whole, even bracketing for the moment its effects on the individual who might be punished. Punishment—especially in the form of incarceration—may not be a rational response to crime, if what we want to do is to reduce crime and improve overall welfare.<sup>78</sup> At the very least, it should be an open question of whether we should sanction people who violate

---

76. RAWLS, *supra* note 7, at 315.

77. I find traces of this view in RAWLS, *supra* note 7, at 241–43, but I will not conduct a textual exegesis here.

78. See, e.g., TODD R. CLEAR, *IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE* ch. 7 (2007).

the law, rather than treat them or forgive them.<sup>79</sup>

I see two major parts to this reconstructed Rawls on punishment. The first, and more important, is to treat punishment in society as something that is not a matter of desert, but a matter of efficiency: we are not trying to fix blame, but trying to fix a problem. How do we decrease crime? How do we organize the scheme of benefits and burdens in order to decrease the possibility that some will be tempted to break the law? One way of doing this will be sanctions, but there are other ways—indeed (and this may be implicit in Rawls all along) a big way will be assuring that people have enough, and the crime will not be a response to need. But we cannot assume that in the well-ordered society *no one* will ever break the law. This is too much of an idealization. Even when people have enough, and have been raised well, there may still be crime and we will have to ask how to approach it in our non-ideal world. Do we say, at that point, that the people who have enough or who aren't deterred by threats of sanctions *deserve* punishment because such acts reveal their “bad character.” Or do we instead take a more holistic view, and say what good is really achieved by punishment in terms of preventing future crimes and reject the idea that punishment is primarily retributive?

The second part to the reconstructed Rawls deals with function of punishment. If we take the Rawls of Part III of *Theory* seriously, I think it suggests that we should view in large part the function of *actual* punishment—as opposed to threatened punishment—as rehabilitative, as opposed to retributive. We should try to show people that their good is in following their sense of justice. We might have an interest in doing this, beyond considerations of efficiency—this is part and parcel of imagining that *we ourselves* when we lift the veil might be subject to punishment. It is also part of the idea that those who are punished may themselves be members of the “least well off,” or at least that punishment could make them so. This thought—that when deciding on what the purpose of punishment is, we should consider that we ourselves may be the object of punishment—that puts a check on the consequentialism of the first part. We would not want to license just *any* treatment of those whom we deem to need sanctions. In fact,

---

79. I also think (and argue in an unpublished essay) that those who are punished too harshly might complain justly about “the strains of commitment” they face in society. In fact, I think many in present-day American society may legitimately make such a complaint. See Flanders, *supra* note 3.

I think it tends to the position that though we may threaten punishment, *actual* punishment may not be justified, not as a matter of consequences, nor even as a matter of how we, looking sympathetically at ourselves pre-the lifting of the veil, would like ourselves to be treated.

It seems to me we can bring both of these steps to a point. Taking a page from Sharon Dolovich, in her magisterial (and lengthy) reconstruction of Rawls on criminal justice,<sup>80</sup> we can see ourselves, from behind the veil as either potential *victims* of crime or potential *objects* of punishment. As potential victims of crimes, our interest is in not being victimized by crime in the first place, so overall in setting up a system of notices and incentives and punishments to prevent this from happening. We should have no first order interest just in seeing people punished.<sup>81</sup> As potential objects of punishment, we should wish to avoid punishment—and the suffering attendant to it—and we should hope that we might be instead objects of *reform*, so that we can better follow the principles of justice. Any interest we might have in purely expiatory suffering we work on satisfying on our own time because from behind the veil we cannot know that we would have such a comprehensive point of view.<sup>82</sup> We cannot, and should not, impute such a controversial view onto our post-veil selves.

The bottom line seems to be this: from either perspective—that of a victim of crime or that of an object of punishment—the idea that we have a system of penal sanctions for *purely* retributive or *purely* denunciatory purposes does not find a place. Other theories fit better for the purposes of a public political conception of justice. Rawls tinkers with jettisoning the idea of punishment as “purely retributive or denunciatory” in the well-ordered society, where we have perfect compliance.<sup>83</sup> But why can it not also be true in a world of less than complete compliance?

---

80. See Dolovich, *supra* note 8.

81. I suppose, privately, we might delight in seeing the wicked suffer; but this is a matter of our personal conception of the good. It is not appropriate for a public political conception (especially since we do not know—pre-veil being lifted—that we actually *will* delight in the wicked suffering).

82. These lines are meant to address the view (given eloquent expression by Michael S. Moore) that if we did wrong, we would *want* to suffer. See Michael S. Moore, *The Moral Worth of Retribution*, in RESPONSIBILITY, CHARACTER, AND THE EMOTIONS: NEW ESSAYS IN MORAL PSYCHOLOGY (Ferdinand Schoeman ed., 1987).

83. RAWLS, *supra* note 7, at 241.

## IV. CONCLUSION

In her book, *Frontiers of Justice*, Martha Nussbaum looks closely at those groups that are excluded from Rawls's social contract theory, mainly for methodological reasons: the sick and the old, members of other nations, animals, etc.<sup>84</sup> Her point is that when we include those groups in a theory of justice, we need to make major revisions to the theory—we cannot leave everything as it was, and just slot these groups in. Rather, the theory requires a fundamental reworking to accommodate these groups. And it *should* accommodate these groups, Nussbaum insists: a good theory of justice cannot just leave certain groups out, or put them off to later.<sup>85</sup> They need to be in at the ground level, something a theory of justice better tell us something about, to be a theory of *justice* worth its salt.

Nussbaum's groups, however, all still work within the law.<sup>86</sup> My paper has considered those who violate the law. Should not a fully worked-out theory of justice tell us how to deal with them? In one way, this can be read as a complaint about ideal theory, a type of theorizing that tells us to disregard those who would break the law *ab initio*. We deal with them later. Of course, Rawls does deal with criminals in his book—but in a by-the-way fashion, and one that is less than appealing. Because he is writing about those who do follow the law, those who do not follow the law look bad: they are those we threaten with sanctions, or who really do deserve to be treated badly. At best, we ignore them.

But what if we included them in the social contract from the beginning? Not to let them dictate the principles of justice, so that raping and looting are somehow permissible. Not this. But we could still see them as fellow citizens in our social union of social unions, just those who have lost their way. We do not have to deal with them with sanctions (unless these are absolutely necessary to prevent more crime overall) nor do we need to exclude them from consideration as the least well off. We can still see what criminals do as bad but not as disqualifying them from treatment as citizens of a well-ordered society.

The structure and methodology of *Theory* moves us away from this type of thinking. Criminals are outside and other.<sup>87</sup> At worst, if they creep into the well ordered society then we pummel them into

---

84. MARTHA C. NUSSBAUM, *FRONTIERS OF JUSTICE* (2006).

85. Nussbaum's overall argument (summarized here) is itself summarized at *id.* at 3–4.

86. With animals, they might not be strictly speaking either inside or outside of the law.

87. See Honig, *supra* note 14.

compliance. Or maybe they simply do not form a part of our theory. The legacy of this treatment in political philosophy is obvious. In much the same way that *Political Liberalism* was thought to cede economic justice to conservatives and only worry about “cultural issues,”<sup>88</sup> so too did *Theory* earlier cede punishment theory to lawyers and politicians.<sup>89</sup> It gave philosophers permission not to think about criminals as part of a theory of justice, and when the prison boom boomed in the 1970s, philosophers were left flat-footed. They still are, for the most part, playing catch-up.<sup>90</sup>

Perhaps what is best in Rawls can be used to fix what Rawls, early on, decided to leave out. Maybe the solution to leaving out criminals is to bring them back in, and think of the distribution of punishment in the same way as we think of the distribution of primary goods: not as a matter of desert, but as a matter of the good of the whole, and of each and every person that is part of that whole.

---

88. The seeds of this objection can be found in Brian Barry’s early review of Rawls, *John Rawls and the Search for Stability* 105 ETHICS 874, 913 (1995), where he argues that in the later Rawls, the “difference principle” may have to be sacrificed to a desire for consensus among plural points of view.

89. Rawls is not alone in this regard; nearly every major book on political philosophy in the late twentieth century says next to nothing about punishment. It was not always this way. See Flanders & Hoskins, *supra* note 2, at 3 (noting that criminal justice is properly a problem of political philosophy).

90. See, e.g., Kwame Anthony Appiah, *What Will Future Generations Condemn Us For?*, WASH. POST, Sept. 26, 2010, at B01 (“We already know that the massive waste of life in our prisons is morally troubling; those who defend the conditions of incarceration usually do so in non-moral terms (citing costs or the administrative difficulty of reforms); and we’re inclined to avert our eyes from the details.”).