

Legitimate Punishment in Liberal Democracy

Sharon Dolovich†

I. Introduction	310
A. Punishment's Legitimacy Problem	310
B. Rawls and Punishment	316
II. Rawls, the Original Position, and the Modeling of Moral Equality	326
A. Criminal Punishment and the Problem of Partial Compliance	326
B. The Original Position and the Veil of Ignorance	329
C. Attributes and Actions: Morally Arbitrary Characteristics and the Distribution of Goods	336
D. Uncertainty and Risk: Maximizing Personal Welfare from Behind the Veil	342

† Acting Professor of Law, UCLA School of Law. I owe my greatest debt to Seana Shiffrin, for her advice, encouragement, insights, and patience through innumerable drafts of this article. I am also grateful to a number of people who provided helpful comments on earlier drafts: Rick Abel, Matt Adler, Taimie Bryant, Devon Carbado, Ann Carlson, Scott Cummings, David Dolinko, Jerry Frug, Stephen Gardbaum, Carole Goldberg, Robert Goldstein, Laura Gomez, Joel Handler, Pamela Hieronymi, Russell Korobkin, Maximo Langer, Gillian Lester, Herb Morris, Steve Munzer, Chris Naticchia, Grant Nelson, Jim Nickel, Tim Kaufman Osborn, Randy Peerenboom, Bill Rubenstein, Mary Sigler, David Sklansky, Kirk Stark, Adam Winkler, Steve Yeazell, Jonathan Zasloff, and Ben Zipursky. Thanks are also due to members of the UCLA Junior Group and to faculty colloquium participants at the Arizona State University College of Law, for extremely helpful discussion; to the members of the Southern California Law & Philosophy Group for early encouragement of this project; and to my colleagues at the Harvard Center for Ethics and the Professions during 1999-2000, who first sounded me out on the themes in this paper and provided helpful guidance. This research was generously supported by the UCLA Academic Senate, the Dean's Fund of the UCLA School of Law, and, in its earliest stages, by Dennis Thompson and the Harvard Center for Ethics and the Professions. Kelly Fernald-Tiberini, Robert Horton, Joseph Schneider and the staff of the Buffalo Criminal Law Review provided excellent editorial assistance. I am particularly grateful to Mikel-Meredith Weidman for her remarkable and extensive editorial and research assistance; the arguments presented herein are much stronger for having faced her critical eye. Finally, I owe a special thanks to Jody Freeman. This article is dedicated to the memory of my father.

1. Gambling under Uncertainty.....	343
2. The Strains of Commitment.....	346
III. Perspectives on Punishment in a Partially Compliant Society.....	350
A. From Ideal Theory to Partial Compliance	350
B. The Priority of Security and Integrity	352
C. Locating Crime Victims and Convicted Offenders Behind the Veil.....	356
1. Criminal Offenders and the Moral Powers..	358
2. The Contingencies of Punishment and Crime	364
a. The Threat of Wrongful Conviction	366
b. Humanity's Imperfections and the Inevitability of Moral Error	368
c. The Arbitrariness of Pressures to Offend—and of the Moral Resources Necessary Always to Resist	369
D. The Rawlsian Model and the Managing of Emotions.....	374
IV. Deriving the Principles.....	378
A. The Lexical Difference Principle and the Deterrent Effect.....	379
B. Punishment in a Partially Compliant Society..	385
1. Punishment of Non-Serious Offenses	386
2. Punishment of Serious Offenses	390
3. Disproportionately Severe Punishment of Serious Offenses	394
4. Disproportionately Severe Punishment of Non-Serious Offenses	399
5. The Parsimony Principle.....	400
6. Determining Evaluative Standards.....	402
C. The Inherent Illegitimacy of Wrongful Conviction.....	404
D. Enumerating the Principles.....	408
E. Incarceration under Inhumane Conditions	409
1. The Problem of Inhumane Punishment.....	409

2. Inhumane Punishment and the Choice among Evils	412
V. From Principles to Policies	419
A. Realizing the Ideal	419
B. The Modified Veil	421
C. The Four-Stage Sequence and the Limits of the Legislative Stage.....	424
D. American Criminal Justice Policy and the Demands of Legitimacy.....	428
1. Obstacles to Careful Deliberation.....	429
2. The Authorization of Illegitimate Practices..	434
VI. Conclusion.....	440

*Those whom we would banish from society or from the human community itself often speak in too faint a voice to be heard above society's demand for punishment.*¹

I. INTRODUCTION

A. *Punishment's Legitimacy Problem*

Any theory of state punishment in a liberal democracy must grapple with the problem of political legitimacy. The punishment of criminal offenders can involve the infliction of extended deprivations of liberty, ongoing hardship and humiliation, and even death. Ordinarily, such treatment would be judged morally wrong² and roundly condemned, yet in the name of criminal justice, it is routinely imposed on members of society by state officials whose authority to act in these ways toward sentenced offenders is generally taken for granted.

Yet should we take the exercise of such power for granted?³ The current scale of incarceration in the United States makes this a particularly appropriate time to raise the legitimacy question in the punishment context. There are now over two million people behind bars in the United States.⁴ This number represents an incarceration rate of

1. *McCleskey v. Kemp*, 481 U.S. 279, 343 (1987) (Brennan, J., dissenting).

2. As Jeffrie Murphy puts it, "[i]f locking human beings in cages or killing them is not a bad way to treat people, it is hard to imagine what would be." Jeffrie G. Murphy, Introduction to Punishment and Rehabilitation 1, 1 (Jeffrie G. Murphy ed., 3d ed. 1995); see also R.A. Duff & D. Garland, Introduction: Thinking about Punishment, in *A Reader on Punishment* 1, 2 (Antony Duff & David Garland eds., 1994) ("[Punishment] is morally problematic because it involves doing things to people that (when not described as 'punishment') seem morally wrong. It is usually wrong to lock people up, to take their money without return, or put them to death.").

3. Judith Shklar, for one, cautions against any such complacency. See Judith N. Shklar, *The Liberalism of Fear*, in *Liberalism and the Moral Life* 21 (Nancy L. Rosenblum ed., 1989). See *infra* note 170 on her "liberalism of fear."

4. See Paige M. Harrison & Allen J. Beck, U.S. Dep't of Justice, *Prisoners in 2002* 1 (2003), <http://www.ojp.usdoj.gov/bjs/pub/pdf/p02.pdf>.

approximately 701 per 100,000 people, up from the 1972 figure of 160 per 100,000.⁵ These figures make America's incarceration rate the highest in the world, larger now even than Russia's.⁶ As one Australian newspaper put it, the current totals make "incarceration the 32nd most populous state in the Union."⁷

Apart from its sheer scale, two other features of the population of America's prisons and jails combine to sharpen the need for inquiry into the legitimacy of prevailing punishment practices. First, there is the racial make-up of the inmate population: it is disproportionately comprised of people of color, African-Americans in particular.⁸ And second, there is the extent of its indigence: America's prisoners are overwhelmingly, disproportionately poor.⁹ Admittedly, viewed in isolation, the race and class position of America's inmate population tells us nothing regarding the legitimacy of the sentences being served.¹⁰

5. See *id.* at 2 (current rate); Marc Mauer, *Race to Incarcerate* 16 (1999) (1972 rate). By way of contrast, the incarceration rate in Europe "hovers at around 90 per 100,000." John Varoli, *Crime & Punishment: The Legacy and Reality of Russia's Prison System*, *Russian Life*, Oct./Nov. 1999, at 37.

6. See Varoli, *supra* note 5, at 37.

7. Mark Riley, *Tough Approach Strikes Out in US*, *Sydney Morning Herald*, Feb. 18, 2000, at 12.

8. Although whites make up 75.1% of the American population, compared with 12.5% Latinos and 12.3% African Americans, U.S. Census Bureau, *Rankings and Comparisons: Population and Housing Tables tbl.1*, U.S. Census 2000, available at <http://www.census.gov/population/www/cen2000/phc-t1.html>, 16% of state and federal prisoners serving sentences longer than one year in 2000 were Latino, and fully 46% of such prisoners were African American. See Harrison & Beck, *supra* note 4, at 11-12. At the end of 2001, 10% of all African American males age twenty-five to twenty-nine were in prison, as compared with 2.9% of Latinos and 1.2% of white males in the same age group. *Id.*

9. Of all state prisoners arrested in 1997, well over half—from 61.4% to 84.9%, depending on education level—earned less than \$2000 in the month before their arrest. Caroline Wolf Harlow, U.S. Dep't of Justice, *Education and Correctional Populations* 10 (2003), <http://www.ojp.usdoj.gov/bjs/pub/pdf/ecp.pdf>. To take another measure of indigence, less than 20% of felony defendants in the country's seventy-five largest counties in 1996 and only a third of felony defendants in federal court [in] 1998 could afford their own attorneys. Caroline Wolf Harlow, U.S. Dep't of Justice, *Defense Counsel in Criminal Cases* 1 (2000), <http://www.ojp.gov/bjs/pub/pdf/dccc.pdf>. [hereinafter Harlow, *Defense Counsel*].

10. I owe this observation to David Garland. See also B. Honig, *Rawls on Politics and Punishment*, 46 *Pol. Res. Q.* 99 (1993).

The fact, however, that a majority of inmates are members of the nation's most disempowered and politically unpopular groups at the very least suggests that some explanation is in order for the current scale of incarceration in the United States.

What, then, is the source of the legitimacy of criminal punishment in a liberal democracy such as ours?¹¹ Does it lie, perhaps, in the acts of the incarcerated themselves? For if people are behind bars, one might argue, they must have broken the law, and as the popular phrasing would have it, "those who do the crime must do the time." This reasoning falls short, however, because it mistakenly assumes that the particular punishment meted out for any offense is inherent in the offense itself. Yet a crime no more dictates the appropriate punishment for its commission than a particular act of misbehavior by a child dictates the necessary parental response to that misbehavior. Rather, in each case, the penalty ultimately imposed represents a normative decision—in the case of crime, a political decision—undertaken by those actors authorized to deploy coercive power under certain conditions. The legitimacy of the specific punishment ultimately imposed for any particular crime, therefore, must lie elsewhere than in the offense itself.

If the particulars of sentencing policies are normative and political, then perhaps the legitimacy of these policies

11. By my use of the term "liberal," I mean to invoke not the currently fashionable meaning of the term as a catch-all contrast to "conservative," but rather the tradition of political theory borne of the Enlightenment challenge to the corporatist, oligarchic, and aristocratic political systems in place across western Europe in the late eighteenth century. See J.G. Merquior, *Liberalism Old and New* 6-7 (1991). This tradition accords moral and political primacy to the individual, elevates individual liberty in its many forms to the highest political value, see *id.*, and measures the legitimacy of political systems by the degree to which they accord sovereignty to the people. For purposes of this article, I therefore assume that a liberal democratic society—or at least a society aspiring to this status—is one with a stated commitment to what I refer to in the text, see *infra* pp. 313-14, as the "baseline" liberal democratic values. And on this definition, the United States, the political life of which is routinely punctuated with the rhetorical invocation of these very values, qualifies as an aspiring liberal democracy.

may simply be found in the political process itself, and in particular in the status of legislators who wrote the laws as duly elected democratic representatives. Voters have, after all, chosen in democratic fashion the representatives whose political commitments best reflect their own preferences. So long as these legislators pass legislation proscribing acts and prescribing their punishment in conformity with established procedural rules, surely the resulting punishments imposed may be safely assumed to be legitimate.

Certainly, the prevailing procedural account accurately identifies as a *descriptive* matter how it is that criminal punishments imposed on convicted offenders are legitimated: in the current system, punishments are viewed as legitimate when they are imposed pursuant to statutes adopted consistent with accepted democratic procedures.¹² As a *normative* matter, however, democratic majoritarianism is a troubling standard on which to claim legitimacy for legislative authorizations of state punishment. For there is nothing inherent in the majoritarian standard to ensure that legislators even fairly consider the interests of all citizens subject to the laws they pass. Indeed, where the targets are politically disenfranchised minorities, the politically powerful need not consider at all the potential harms their preferred policies might impose on members of these powerless groups.

To determine the normative constraints placed on the state's power to punish in a liberal democracy, it may be more promising to look to the normative commitments of such a system. If the idea of a liberal democracy means

12. Technically, in the current system, there is a further step that may also be required to affirm the legitimacy of state punishments: the determination by a court that the criminal sentence does not violate the Eighth Amendment's prohibition on cruel and unusual punishment. In practice, however, the Supreme Court has so narrowly construed this constitutional provision that judicial review of any criminal sentence short of the death penalty effectively functions as a rubber stamp of the sentence imposed. See, e.g., *Rummel v. Estelle*, 445 U.S. 263, 272 (1980) (holding that a sentence of life in prison did not violate the Eighth Amendment when imposed on a recidivist who obtained \$120.75 through false pretenses, and noting that "[o]utside the context of capital punishment, successful challenges to the proportionality of particular sentences have been exceedingly rare").

anything, it means a commitment to what we can think of as the "baseline" liberal democratic values: individual liberty, dignity, and bodily integrity; limited government; the primacy and sovereignty of the individual; and the entitlement of all citizens to equal consideration and respect.¹³ And although others may derive a different conclusion as to what consistency with these values would require, as I see it, if state power in any of its forms is to be exercised consistently with these baseline values, it cannot be of a form or scope in which citizens would merely acquiesce, whether out of fear or because they were out-voted or because they had no choice. Rather, they must in some sense accept the terms of its exercise as reflecting what Rawls has called the "fair terms of social cooperation."¹⁴ If the exercise of state power in a liberal democracy is to be legitimate, that is, it must be justifiable in terms that all members of society subject to that power would accept as just and fair.¹⁵ This imperative is particularly acute in the context of criminal punishment, for of all the manifestations of power the state exercises against its own citizens, the punishment of convicted offenders is the most intrusive, and the most severe.

It may be hard at first to see how this standard could yield the terms of legitimate punishment. For wouldn't any citizens, finding themselves convicted of crimes and facing punishment, simply withhold their agreement as to the just and fair nature of the contemplated penalty, whatever its character? Plainly, if this standard is to ground the normative legitimacy of state punishment, there must be some way for citizens to maintain the idea of meaningful agreement while abstracting consideration of the particular details of their individual lives.

The work of John Rawls offers a way to strike this balance. Rawls argues that political power is legitimately exercised by the state over its citizens only when it is

13. See *supra* note 11.

14. John Rawls, *A Theory of Justice* 21 (1971) [hereinafter TJ].

15. In conceiving of political legitimacy in this way, I am of course greatly indebted to Rawls.

exercised on the basis of a collective agreement "the essentials of which all citizens may reasonably be expected to endorse" under fair deliberative conditions.¹⁶ As he sees it, such fair deliberative conditions are those that allow consideration of the terms of state power from a "suitably general point of view,"¹⁷ which he defines as the perspective from which no participant to the deliberative process knows anything about the particulars of his or her own personal identity or social position.¹⁸ On Rawls's view, that is, if state power is to be legitimate, agreement as to the terms of its exercise must come from citizens who do not know the first thing about their own situation and who must therefore accord due consideration to the perspectives of all members of society.

In *A Theory of Justice*, Rawls introduces the concept of the original position, with its veil of ignorance, as a way to model deliberations that would yield the content of such an agreement.¹⁹ Rawls himself does not apply this model to the problem of punishment. Yet if his conception of the "suitably general point of view" is a fair characterization of the appropriate perspective for judging the legitimacy of state power, as I believe it is, it would seem worthwhile to put his model to work in this way. In this article, I do just that. My aim is to identify principles of punishment which we would all accept as just and fair if we were to find ourselves behind a veil of ignorance suitably framed for a

16. John Rawls, *Political Liberalism* 217 (1993) [hereinafter PL]. This is the idea behind what Rawls refers to as the "liberal principle of legitimacy," that "our exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in light of principles and ideals acceptable to them as reasonable and rational." *Id.*

17. TJ, *supra* note 14, at 304.

18. See *id.* at 136-42.

19. Throughout this article, I draw primarily on the arguments Rawls develops in the original 1971 edition of *A Theory of Justice*. I refer to other of his writings when the discussions therein help to clarify concepts or arguments presented in that work. When the references I make to later works reflect aspects of his views on which Rawls changed his position over time, I indicate this fact in the notes. Otherwise, the reader should assume that references to later works are consistent with the argument Rawls presents in the original edition of *A Theory of Justice*.

social context in which the problem of punishment is salient. It is punishment imposed consistent with such principles, I argue, that constitutes legitimate punishment in liberal democracy.²⁰

B. Rawls and Punishment

This application of Rawls's model may, I realize, strike some readers as wrongheaded. For this model in a sense demands strict impartiality among the interests of all members of society, including the interests of potential criminals. And, some readers might insist, let's face it: we don't *want* to be impartial vis-à-vis the interests of criminals. They are, after all, criminals. Some of them have done awful things to others and all of them have broken the law, and they have thus forfeited their right to this kind of consideration.

Certainly, those individuals who have violated the security and integrity of their fellow citizens have thereby done serious wrong and deserve our condemnation and scorn. Indeed, the notion that criminals have by their criminal actions forfeited their right to the same measure of social goods afforded the law-abiding is fully consistent both with Rawls's model and with the present project. The approach I take parts company with this forfeiture view, however, to the extent that it suggests that, by their actions, criminal offenders not only forfeit their right to equal treatment, but also place themselves outside the circle of those who are entitled to full consideration as subjects of justice. For it is a basic assumption of the argument I develop—and indeed, of liberal democracy itself—that all members of society are moral equals, entitled to due consideration and respect as fellow human beings and fellow citizens. From this assumption, it does not follow that all citizens are entitled to equal *treatment*. To the contrary, by their actions, individuals may forfeit

20. Thus, when I refer to "legitimate punishment" herein, I mean punishment imposed consistently with the principles that all would agree to be just and fair when considered from the impartial conditions of the original position.

certain goods that other citizens enjoy. But on the theory of liberalism I adopt here, forfeiture in this sense does not negate an individual's moral status:²¹ he or she is still a subject of justice, entitled to consideration as such. The argument I offer is thus first and foremost directed at those readers who accept this baseline premise. My aim is to determine what a theory of punishment would look like for a liberal theory that subscribes to this conception of prior moral equality²² and does not condition such prior moral status on citizens' good behavior.

Given the deeply emotional character of any discussion of crime, to ensure that due consideration is extended even to potential criminals is a difficult task. Rawls's framework is promising for our purposes precisely *because* it demands that we consider with some measure of dispassion even the interests of potential criminal offenders. That we might be inclined to resist applying this framework because it demands this limited measure of emotional detachment only goes to show how necessary it is to impose such dispassion, artificially if necessary, if we are to hope to end up with principles that all members of society may freely affirm as just and fair.

There is, however, a further objection to the present project that might be raised even by those who share its normative commitments—that even if we were inclined to bring Rawls's framework to bear on the problem of punishment, the effort would be doomed from the start. For Rawls's model yields its results by denying the parties knowledge of all morally arbitrary attributes and personal

21. Indeed, some have argued that it is by punishing those who have transgressed that we are manifesting our respect for their equal moral status. See, e.g., R.A. Duff, *Trials and Punishments* 52 (1986); Herbert Morris, *Persons and Punishment*, 52 *Monist* 74-75 (1968), reprinted in *Punishment and Rehabilitation*, *supra* note 2, at 74-93.

22. By "prior moral equality," I mean to indicate the moral status to which members of society are entitled simply by virtue of their shared humanity. Certainly, if someone commits a morally culpable act that violates the security and integrity of another, in an important sense he or she is no longer viewed by society as entitled to equal treatment. Such moral culpability, and the unequal treatment to which it may give rise, does not, however, imply that the guilty are any less deserving of moral consideration as subjects of justice.

particulars. And, it might be argued, whether or not one faces punishment as a convicted offender is not morally arbitrary. It would instead hinge on the target's having actually undertaken criminal actions, actions that are not morally arbitrary individual attributes hidden by the veil of ignorance,²³ but instead the product of moral choices for which the actor is properly held responsible. The parties would therefore not face uncertainty behind the veil as to whether they themselves could wind up targets of state punishment. And if the parties do not face uncertainty on this score behind the veil, they will not in their deliberations be moved to protect themselves from the possible future experience of state punishment. And, for this reason, they will not accord any consideration in their deliberations to the perspective and interests of potential criminal offenders.

Granted, Rawls does draw a distinction between attributes and actions, one on which attributes but not actions are understood to be morally arbitrary for purposes of deliberations behind the veil. On his framework, once the veil is lifted and the parties are understood to enter a society governed by the principles of justice, the parties are no longer viewed simply as inactive bearers of morally arbitrary attributes. They are now citizens in society, moral actors who are held to be responsible for the effects of their actions on their distributive shares—and on the distributive shares of others.²⁴

23. On the veil of ignorance and its role in Rawls's framework, see *infra* part II.B.

24. One might object that our actions are to some extent inseparable from our attributes, and thus to the extent that we are punished for our actions, such treatment is traceable to the product of morally arbitrary contingencies. There is surely something to this view. As we will see, however, Rawls is not—as is sometimes supposed—committed to the luck egalitarian position that the purpose of a theory of justice is to compensate for all the negative effects of unchosen circumstances. See *infra* part II.C. His view is the very different one that we respect each other as moral equals when we govern our basic structure on the basis of principles of justice we would select in an initial position of moral equality. The parties under such conditions would reject the alternative social vision offered by principles that compensate in the distribution of society's goods (including the goods of security and integrity, see *infra* part III.B) for any

To say that parties as citizens will ultimately be held responsible for their actions, however, is not the same as saying that, considering the question from behind the veil, the parties would have no grounds for uncertainty as to whether, once the veil is lifted, they could wind up convicted offenders facing punishment. Rather, the crucial question for parties trying to maximize their own interests is whether, despite their ignorance of the nature of their own (morally arbitrary) attributes, they could nevertheless be fully confident that they would always be in sufficient control over their (morally relevant) actions to guarantee that they would always be able to avoid any criminal actions for which they would be held fully responsible and punished, perhaps severely. And, given the conditions of partial compliance (which for our purposes we must assume),²⁵ the answer to this question must be no, for three reasons.²⁶ First, the danger of wrongful convictions in a partially compliant society means that even innocent people could find themselves facing criminal punishment once the veil is lifted. Second, although all citizens in a partially compliant society are stipulated to have the basic moral powers to the requisite minimum degree, the parties still know that they are human beings, with all the qualities of impulsiveness, bad judgment, proneness to error, and other limitations this status entails. And third and finally, an unjust distribution of society's goods in a partially compliant society means that citizens will differ dramatically in terms of both the pressures and temptations they face to offend against others, and the economic and moral resources with which they are equipped to resist such pressures and temptations. As I argue in part III below, these three factors in combination

negative consequences traceable to morally arbitrary attributes. For this reason, the arguable influence of such morally arbitrary contingencies on the actions one undertakes is not compensated for in the principles of justice ultimately selected.

25. For a description of the conditions of partial compliance, see *infra* part III.A. On the concept of partial compliance in contrast with Rawlsian ideal theory, see *infra* p. 324 and part II.A.

26. See *infra* part III.C.2 for fuller argument supporting this claim.

are sufficient²⁷ to create conditions of uncertainty for the parties in the original position²⁸ as to whether they themselves would turn out to be targets of punishment in a partially compliant society—and would, I argue, lead them to consider the perspective of convicted offenders along with that of potential victims when choosing among proposed principles of punishment.

At the same time, as we will see, the parties' uncertainty as to whether they might find themselves facing criminal punishment once the veil is lifted in no way leads them to the conclusion that criminal offenders should not pay a price for their crimes. Nor do the parties conclude that criminals and their victims are somehow "on a moral par" or deserve equal treatment. Despite the parties' readiness (for their own benefit) to consider the perspective of the targets of punishment when selecting the principles of justice, the theory of punishment that emerges from the parties' deliberations does hold offenders responsible for their actions, condemn their crimes as moral wrongs, and authorize punishment, at times severe, for offenses committed against others. As I demonstrate below, these conclusions are fully consistent with a theory that accords all members of society—potential criminals included—the consideration and respect necessary to ensure that any state power exercised over them is justified in terms that they could accept as just and fair.

That the theory of punishment to emerge from the Rawlsian framework condemns criminal actions and holds criminal offenders responsible for their crimes should come as no surprise. Rawls's theoretical model, after all, is self-consciously derived from Kant,²⁹ whose view of punishment

27. Indeed, as I argue below, see *infra* note 175, the first two factors taken alone are sufficient to make the case. The third—the most contentious of the three—only serves to reinforce the point.

28. On Rawls's concept of the original position, see *infra* part II.B.

29. See, e.g., TJ, *supra* note 14, at 140-41 ("The notion of the veil of ignorance is implicit, I think, in Kant's ethics."); *id.* at 179 ("[T]he principles of justice manifest in the basic structure of society men's desire to treat one another not as means only but as ends in themselves. I cannot examine Kant's view here. Instead I shall freely interpret it in the light of the contract doctrine."); *id.* at 256

is famously retributivist.³⁰ At the same time, the theory of punishment I develop below is by no means purely retributivist, and indeed may even appear to some readers to be consistent, not with retributivism, but with a straightforwardly utilitarian focus on punishment's deterrence function. In fact, the theory of punishment that flows from a Rawlsian analysis eludes categorization as either purely retributivist or purely utilitarian. Deterrence calculations do drive the sentencing determinations to be made consistent with the principles of punishment that, I argue, would be chosen in the original position. Such punishments are nonetheless imposed with a backdrop of notions and limits that are far more consistent with a retributivist view than with a utilitarian one. True, on this model, the fact that one is found in a moral sense to deserve a certain quantum of punishment is not alone a *sufficient* basis for the state legitimately to impose punishment in that measure.³¹ But as we will see, moral desert is nonetheless a *necessary* condition for legitimate punishment, a feature which alone distinguishes the theory here presented from a pure deterrence theory, notwithstanding the utilitarian cast of the principles. If there is an analogue in traditional justifications for punishment to the Rawlsian theory I develop, it is a mixed theory, one with elements of both retribution and deterrence.³²

("The original position may be viewed . . . as a procedural interpretation of Kant's conception of autonomy and the categorical imperative.").

30. See Immanuel Kant, *The Philosophy of Law* 195 (W. Hastie trans., 1974) (1887) ("Juridical Punishment can never be administered merely as a means for promoting another Good either with regard to the Criminal himself or to Civil Society, but must in all cases be imposed only because the individual on whom it is inflicted *has committed a Crime*."). See also id. at 198 ("Even if a Civil Society resolved to dissolve itself with the consent of all its members . . . the last Murderer lying in the prison ought to be executed before the resolution was carried out . . . in order that every one may realize the desert of his deeds . . .").

31. The Rawlsian theory I develop here is thus incompatible with a pure retributivist view, on which "[m]oral culpability ('desert') is . . . both a sufficient as well as a necessary condition of liability to punitive sanctions." Michael S. Moore, *The Moral Worth of Retribution*, in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* 179, 181-82 (Ferdinand Schoeman ed., 1987).

32. That this is so should also come as no surprise to those familiar with

Although Rawls himself never addressed the punishment question directly, at many points in his writing he indicates his expectation that such an analysis would follow from the ideal theory he advances. Yet strangely, despite the voluminous scholarly attention paid both to Rawls's ideas³³ and to the goal of finding a satisfactory moral justification for punishment, very little has been written attempting to resolve the problem of punishment using Rawls's model.³⁴ Nor have scholars of punishment—

Rawls's work; as with Rawls's own theory of justice as fairness, the theory of legitimate punishment developed herein derives from a model constructed of free and equal citizens possessed of the two moral powers, who are motivated to maximize their own shares of society's goods behind the veil.

33. See Anthony Simon Laden, *The House that Jack Built: Thirty Years of Reading Rawls*, 113 *Ethics* 367, 367 (2003) ("Over the course of the last half-century, roughly three thousand articles that discuss the work of John Rawls have been published in journals of philosophy, law, economics, political science, and related fields.").

34. Of those who *have* addressed the problem of punishment using Rawls's model, none has attempted, as I do, systematically to derive from the terms of the original position principles to constrain the state's exercise of its punitive power. David Hoekema comes closest to approaching the question in this way. See David A. Hoekema, *The Right to Punish and the Right to Be Punished*, in John Rawls' *Theory of Justice: An Introduction* 239, 259 (H. Gene Blocker & Elizabeth H. Smith eds., 1980). But because Hoekema starts with the traditional moral justifications for punishment and not with the nature of the parties' interests in enabling and also limiting the state's punitive power, the answers he derives do not speak to the problem of constraining state power as I understand it, and are consequently less broadly applicable than those I seek here. The same may be said for Samuel Donnelly. See Samuel J.M. Donnelly, *The Goals of Criminal Punishment: A Rawlsian Theory (Ultimately Grounded in Multiple Views Concerning Human Dignity)*, 41 *Syracuse L. Rev.* 741 (1990). But see Jeffrie G. Murphy, *Retributivism, Moral Education, and the Liberal State*, 4 *Crim. Just. Ethics* 3, 4, 8 (1985) (suggesting the potential of political theory in general, and Rawlsian social contract theory in particular, to provide the basis for a theory of punishment addressing "the problems of the nature and justification of the state and its coercive power"). On other aspects of the question of Rawls and punishment, see Honig, *supra* note 10; Samuel Scheffler, *Justice and Desert in Liberal Theory*, 88 *Cal. L. Rev.* 965, 978 (2000).

As far as I know, Thomas Pogge alone has addressed the relationship between Rawls and punishment through the lens of state power, although he takes a somewhat different approach than I do here. See Thomas W. Pogge, *Three Problems with Contractarian-Consequentialist Ways of Assessing Social Institutions*, 12 *Soc. Phil. & Pol'y* 241 (1995). In the course of developing a critique of what he calls "contractarian-consequentialist" theories (a category that includes Rawls's justice as fairness), Pogge presents a set of challenges for any attempt to apply the Rawlsian model to the problem of punishment. Although his

even those explicitly committed to justifying state punishment in terms of central liberal values³⁵—paid much attention to the broader questions with which this article is concerned, questions relating to the coercive nature of state power and to the consistency of state punishment with the justificatory standards of particular state forms.³⁶ By

analysis is thought-provoking, I find his arguments to be ultimately unpersuasive, for they fail to grapple with the full complexities of Rawls's theory. Most notably, Pogge treats the parties' concern with maximizing their positions as the only constraint on the model, when in fact this focus on consequences cannot be understood independently of the broader (non-consequentialist) normative commitments of the framework as a whole. Pogge's arguments also collapse two crucial distinctions on which Rawls's theory is deliberately and coherently constructed: (1) the distinction between ideal theory and the problems of partial compliance; and (2) the distinction between the selection of general principles of justice in the original position, and the translation of these general principles into policy at the third legislative stage. In this article, I trace the aspects of Rawls's theory that would—as Rawls himself anticipated—allow its application to the problems of partial compliance. In so doing, I hope to illustrate both the necessity of preserving the distinctions Pogge's analysis collapses, and the potential of Rawls's theory to address the problem of punishment notwithstanding the admittedly difficult puzzles Pogge presents.

35. A commitment to liberal values has informed the work of many influential scholars of punishment. See, e.g., Joel Feinberg, *The Moral Limits of the Criminal Law* (1984-1988); Jean Hampton, *Retribution and the Liberal State*, 5 *J. Contemp. Legal Issues* 117 (1994); Duff, *supra* note 21; Herbert Morris, *A Paternalistic Theory of Punishment*, 18 *Am. Phil. Q.* 263 (1981), reprinted in *Punishment and Rehabilitation*, *supra* note 2, at 154; Warren Quinn, *The Right to Threaten and the Right to Punish*, 14 *Phil. & Pub. Aff.* 327 (1985). Cf. Nicola Lacey, *State Punishment: Political Principles and Community Values* (1988) (endorsing a communitarian approach to punishment).

36. See Michael Philips, *The Justification of Punishment and the Justification of Political Authority*, 5 *Law & Phil.* 393, 394 (1986) ("[M]ost philosophical discussions of punishment proceed as if the justification of punishment can be understood independently of political philosophy."); Murphy, *supra* note 34, at 3 (arguing that traditional approaches leave "two deep questions" unanswered: "(1) Is it the legitimate business of the state to pursue these wonderful goals (or some of them)? (2) Even if it is the legitimate business of the state to pursue these goals, (or some of them), is it proper for the state to choose the *criminal law* as the appropriate means?").

This lack of attention by philosophers of punishment to these questions is an artifact of the extent to which the punishment debate has been dominated by moral philosophers, while political theorists for the most part have been disinclined to participate. As a consequence, this debate has focused primarily on questions relating to the traditional moral justifications for punishment—deterrence, retribution, rehabilitation, and so on—to the exclusion of questions bearing on state power, which I take as central. Recent work suggests that political theorists have begun to remedy this past neglect. See, e.g., *Punishment*

providing the sketch of a Rawlsian theory of punishment and thereby developing an explicitly liberal theory of state punishment, this article thus fills two notable gaps in the literature.

In addition, in a somewhat different vein, I offer with this article a model for adapting Rawls's methodological framework to problems of partial compliance. Rawls assumes that his framework applies only to questions of what he calls "ideal theory," in which "everyone is presumed to act justly and do his part in upholding just institutions." He recognizes that the problems of what he calls "partial compliance theory"—that is, the problems of the real world—are "the pressing and urgent matters," and indeed he views ideal theory as the necessary starting point to derive solutions to them.³⁷ But he says little about how one is to go about moving from ideal theory to addressing problems of partial compliance.

The difficulty with applying Rawls's framework to problems of partial compliance is that the assumptions of the well-ordered society on which it relies are plainly at odds with the functioning of our own society. In this article, however, I demonstrate that it is possible to apply the Rawlsian model to non-ideal conditions while also preserving the power of the basic deliberative framework to ensure due consideration of the perspectives of all members of society—including the non-compliant. In this way, I show that the Rawlsian model is capable of yielding principles citizens may be expected to endorse under fair deliberative conditions—the hallmark, in Rawls's view, of political legitimacy—even in the context of the messy problems of our actual political world.

The argument of this article proceeds as follows. In part II, I provide a brief overview of Rawls's framework and of his central operating concepts, an understanding of which will be necessary to the argument that follows. I also offer in this part an account of the veil of ignorance as

and Political Theory (Matt Matravers ed., 1999). This article is intended as a contribution to that effort.

37. TJ, *supra* note 14, at 8-9.

obscuring attributes but not actions—an account that is central to the possibility of applying Rawls's model to conditions of partial compliance and which has not, to my knowledge, been heretofore advanced. In part III, I address the problem of rendering Rawls's framework applicable to the problem of punishment in a non-ideal world, identifying three features of his well-ordered society that would need modification for the model to be put to this purpose. The primary aim of this part is to determine the perspective with which the parties, operating behind the veil of ignorance, would approach the task of selecting the principles of punishment to govern society. Armed with this understanding, I turn in part IV to the content of the principles of just punishment the parties would select. Although criminal punishment in contemporary liberal democracies comes in many forms, both for simplicity's sake and because incarceration for a term of years has been the punishment of choice in the United States for much of the nation's history to the present day, I restrict the analysis to consideration of the punishment of incarceration.³⁸ In this part, focusing exclusively on the punishment of incarceration, I identify a number of principles of punishment to which the parties would agree. As we will see, each of the principles ultimately selected proves a variant on what I call the "parsimony principle,"³⁹ the basic idea of which is that the punishment of convicted offenders must be no more severe than necessary to yield

38. In restricting the meaning of punishment in this way, I do not intend to suggest that, were other options open to them, the parties in the original position would necessarily light on incarceration as the sole available punishment, or even as the punishment of choice. To the contrary, it seems very likely that parties confronted with the task of identifying appropriate forms of criminal punishment behind the veil would endorse a wider range of sentencing alternatives than that which I allow. But if our analysis is to equip us to judge the legitimacy of the bulk of the criminal sentences being imposed and served in the United States today—which, as I have already indicated, is a central motivating aim of this project—this restriction is a necessary one.

39. I borrow this term from Braithwaite and Pettit. See John Braithwaite & Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* 87 (1990). See also Jeremy Bentham, *Introduction to the Principles and Morals of Legislation* (W. Pickering 1823).

an appreciable deterrent effect on the commission of serious offenses.

In yielding the principles of punishment to which the parties would agree in the original position, part IV represents the heart of the argument. For if this framework adequately captures our intuitions regarding the fair terms of deliberation over the exercise of state power, these principles would represent the appropriate normative constraints on the state's exercise of its power to punish criminal offenders in a liberal democracy. And it is therefore these principles against which we should measure the legitimacy of particular criminal justice policies and the specific sentences they authorize. If, however, our goal is to use these principles to derive the terms of legitimate criminal justice policies, we must consider a further step: determining how these principles are to be translated into policies applicable to the real world. In part V, I therefore consider the implications of the theoretical framework developed in parts III and IV for this policymaking process, and show that the principles derived in part IV would provide the basis for questioning the legitimacy of a number of prevailing criminal justice policies. Although I recognize that the principles I outline and the account of the legislative process I develop represent a utopian vision, I nonetheless argue that this vision accurately characterizes the obligations legislators owe their constituents even in the distinctly non-utopian context of contemporary American society.

II. RAWLS, THE ORIGINAL POSITION, AND THE MODELING OF MORAL EQUALITY

A. *Criminal Punishment and the Problem of Partial Compliance*

In his work, Rawls does not concern himself directly with the problem of punishment.⁴⁰ Instead, his concern is

40. The problem of punishment surfaces occasionally in *A Theory of Justice*,

with distributive justice, his goal being to identify principles to guide a just distribution of society's "primary goods," those "various social conditions and all-purpose means" that all citizens would need to develop and grow as moral beings and to pursue their own conception of the good.⁴¹ To this end, Rawls remains in his analysis at the level of "ideal theory," assuming a world of "strict compliance," in which institutions are just and "[e]veryone is presumed to act justly and to do his part in upholding [these] just institutions."⁴² In such a world, there would be little crime, and penal sanctions would only be necessary to assure all members of society that all others too will do their part and respect the laws.⁴³

The world of strict compliance is not, of course, our world. To determine meaningful principles of just punishment, we must assume a different social context, one in which neither individuals nor institutions are fully

generally as a point of contrast to the treatment legitimately expected of law-abiding citizens in a well-ordered society. See TJ, *supra* note 14, at 314-15, 575-76. In an early essay, Rawls weighs in on the long-standing debate over the appropriate justification for the institution of punishment (retribution or deterrence?), but again, he does so to illustrate a more general theoretical point. In this case, his point is that it is possible to distinguish between the justification of a rule or practice and the justification "of a particular action falling under it." John Rawls, *Two Concepts of Rules* (1955), reprinted in John Rawls: *Collected Papers* 20, 21 (Samuel Freeman ed., 1999).

41. John Rawls, *Justice as Fairness: A Restatement* 57 (Erin Kelly ed., 2001) [hereinafter JF]. As Kymlicka explains, these are the goods that each person would "want or need to enable [them] to lead a good life." Will Kymlicka, *Contemporary Political Philosophy: An Introduction* 63 (1990). Rawls originally described the primary goods as "things which it is supposed a rational man wants whatever else he wants." TJ, *supra* note 14, at 92. He eventually revised this description, coming to describe primary goods as "various social conditions and all-purpose means that are generally necessary to enable citizens adequately to develop and fully exercise their two moral powers, and to pursue their determinate conceptions of the good." JF, *supra*, at 57; see also PL, *supra* note 16, at 75-76. The list of primary goods has remained as it was enumerated in *A Theory of Justice*, and includes rights and liberties, income and wealth, powers and opportunities, and the social bases of self-respect. See TJ, *supra* note 14, at 92.

42. TJ, *supra* note 14, at 8-9.

43. Rawls calls this the "assurance problem." *Id.* at 315; see also *id.* at 240 (explaining the role authorized penal sanctions play in resolving the assurance problem).

compliant with the demands of justice. In this more familiar world of partial compliance, the reach of a criminal justice system will be necessarily much broader—and its role more urgent.

Rawls is careful to distinguish his own aims, for which he assumes the strict compliance of a well-ordered society, from the goals of what he calls “partial compliance theory,”⁴⁴ which “studies the principles that govern how we are to deal with injustice.”⁴⁵ Indeed, Rawls appears to believe that problems arising within the scope of partial compliance theory would require a different framework for their resolution than the one he applies to the task of identifying principles of just distribution for a strictly compliant society.⁴⁶ But as I demonstrate in this article, the framework Rawls applies to identifying the terms of a just distribution in an ideal world—the now-familiar “original position” with its “veil of ignorance”—is equally effective when applied to the problems of partial compliance, including the particular problem that concerns us here.⁴⁷ As we will see, a suitably modified version of Rawls’s methodology is appropriate for our purposes precisely because it allows us to focus attention on the concerns criminal punishment is intended to address while screening out those considerations that, though irrelevant, tend nonetheless to drive popular thinking and policymaking on this issue. Moreover, because it is designed to accord equal consideration and respect to all members of society, Rawls’s framework ensures that the

44. *Id.* at 8; see also *id.* at 315 (“The question of criminal justice belongs for the most part to partial compliance theory whereas the account of distributive shares belongs to strict compliance theory and so to the consideration of the ideal scheme.”).

45. *Id.* at 8.

46. Rawls assumes that addressing the problems of partial compliance requires full knowledge of the particular society in which the problems arise, a scope of knowledge he denies to the parties on his framework. See *infra* part II.B. He also assumes that partial compliance theory is necessarily dependent on a successful resolution, under the deliberative terms he suggests, of the problem of distributive justice under ideal conditions. See TJ, *supra* note 14, at 245-46.

47. See TJ, *supra* note 14, at 8-9 (distinguishing consideration of “strict compliance” from problems of “partial compliance theory”).

conclusions reached as to the nature of just punishment are those that all members of society could freely accept.

In the sections that follow, I provide a brief explanation of this framework and the nature of the deliberative process that flows from it, before moving on, in parts III and IV, to the central project of this article—considering the principles of just punishment that a suitably modified version of Rawls's framework would yield.

B. The Original Position and the Veil of Ignorance

In *A Theory of Justice*, Rawls identifies two principles on the basis of which, he believes, a just distribution of social goods may be effected.⁴⁸ But more than this, he seeks to show that these two principles, to which he refers collectively as “justice as fairness,” are those that all members of society “can publicly endorse in the light of their own reason.”⁴⁹ This broadly inclusive standard of public justification derives from what Rawls calls the “liberal principle of legitimacy.”⁵⁰ According to this principle, because political power in a liberal democracy is exercised over citizens in the name of the people themselves, if this exercise of power is to be legitimate, it

48. *Justice as Fairness* contains the most recent statement of Rawls's two principles of justice:

(a) Each person has the same inalienable claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all; and

(b) Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least-advantaged members of society (the difference principle).

JF, *supra* note 41, at 42-43. As Rawls explains it, “the first principle is prior to the second; also, in the second principle fair equality of opportunity is prior to the difference principle. This priority means that in applying a principle (or checking it against test cases) we assume that the prior principles are fully satisfied.” *Id.* at 43.

49. *Id.* at 91.

50. PL, *supra* note 16, at 217.

must be justifiable in terms all society's members could reasonably be expected to accept—explicable, that is, in light of “principles and ideals acceptable to [the] common human reason” of all members of society.⁵¹

For Rawls, this demand for broad public justification does not amount to a requirement that all members of society actually consider and agree to the principles of justice to govern the institutions of the basic structure. There is no expectation, or hope, of a society-wide referendum.⁵² Instead, what Rawls seeks are principles that all citizens could reasonably be expected to endorse from “a suitably general point of view”⁵³—a point of view in which all unfair bargaining advantages are eliminated and citizens come together in a situation of mutual consideration and respect.

What are the defining features of this “suitably general point of view”? This question brings us to the “original position,”⁵⁴ the thought experiment⁵⁵ that is Rawls's version of the social contract.⁵⁶ This hypothetical place to which participants are said to retreat to deliberate over questions of justice is intended to represent the “suitably general” conditions that model our considered convictions over the fair terms of such deliberation. The most familiar of these conditions is what Rawls labels the “veil of ignorance.”⁵⁷ When the parties to the original

51. *Id.* at 137; see also JF, *supra* note 41, at 90-91 (explaining that “while political power is always coercive—backed by the government’s monopoly of legal force,” in a liberal democratic regime, it is also the power of the people, “that is, the power of free and equal citizens” as a collective body).

52. To the contrary, in Rawls's view, it is the unequal distribution of bargaining power informing the political process as it actually functions that gives rise to the illegitimate exercise of coercive political power in the first place.

53. TJ, *supra* note 14, at 304.

54. PL, *supra* note 16, at 24. On the original position, see generally TJ, *supra* note 14, at 136-46; PL, *supra*, at 22-28; JF, *supra* note 41, at 14-18; and Kymlicka, *supra* note 41, at 58-70.

55. JF, *supra* note 41, at 17 (describing the original position as “a device of representation, or, alternatively, a thought-experiment for the purpose of public- and self-clarification”).

56. See PL, *supra* note 16, at 23 (“Justice as fairness recasts the doctrine of the social contract . . .”).

57. TJ, *supra* note 14, at 136.

position deliberate, they are said to do so behind this veil. Although the veil allows the parties to know "the general facts about human society," including "the basis of social organization and the laws of human psychology,"⁵⁸ it screens out any knowledge of the parties' own personal characteristics, social or economic status, race or gender,⁵⁹ particular conception of the good, or any other distinguishing features of their own particular identity.⁶⁰ Why impose this restriction? Because these features, although arbitrary from a moral point of view,⁶¹ nonetheless greatly influence the outcome of public deliberation over political questions, frequently ensuring that those who enjoy greater social and economic power prevail over those who do not.⁶² And it is, in Rawls's estimation, one of our considered convictions—one of the "fixed points" in our "shared fund of implicitly recognized basic ideas and principles"⁶³—that "the fact that we occupy a particular social position is not a good reason for us to propose, or to expect others to accept, a conception of

58. Id. at 137.

59. Rawls did not initially include race or gender in the set of identity positions the veil of ignorance obscures. See TJ, supra note 14, at 137. The inclusion of these categories in this set is, however, entirely consistent with the point of the construct, and in both *Political Liberalism* and *Justice as Fairness*, Rawls makes their inclusion explicit. See PL, supra note 16, at 24-25; JF, supra note 41, at 15. I owe this observation to Seana Shiffrin. See Seana Shiffrin, *Race, Labor and the Fair Equality of Opportunity Principle*, 72 *Fordham L. Rev.* (forthcoming, April 2004).

60. See JF, supra note 41, at 15:

In the original position, the parties are not allowed to know the social positions or the particular comprehensive doctrines of the persons they represent. They also do not know persons' race and ethnic group, sex, or various native endowments such as strength and intelligence, all within the normal range. We express these limits on information figuratively by saying the parties are behind a veil of ignorance.

See also TJ, supra note 14, at 137; PL, supra note 16, at 24-25.

61. See TJ, supra note 14, at 311 ("[I]t is one of the fixed points of our moral judgments that no one deserves his place in the distribution of natural assets any more than he deserves his initial starting place in society.").

62. See TJ, supra note 14, at 141 ("If a knowledge of particulars is allowed, then the outcome is biased by arbitrary contingencies.").

63. PL, supra note 16, at 8.

justice that favors those in this position."⁶⁴ Rawls therefore proposes the veil of ignorance. For if deliberations in the original position are to be fair, we "must eliminate the [morally arbitrary] bargaining advantages that inevitably arise over time within any society as a result of cumulative social and historical tendencies."⁶⁵

Behind the veil, the parties are "symmetrically situated,"⁶⁶ a feature that directly derives, Rawls explains, from the "considered conviction that in matters of basic political justice citizens are [free and] equal in all relevant respects"⁶⁷ For Rawls, free and equal citizens are those who possess "to a sufficient degree"⁶⁸ what he calls the "two moral powers":⁶⁹ the capacity to develop, revise, and pursue their own conception of the good,⁷⁰ and the "capacity for a sense of justice," that is, "to understand, to apply, and to act from . . . the principles of political justice that specify the fair terms of social cooperation."⁷¹ This view of the liberal subject is deeply, even radically, egalitarian.⁷² For to qualify for "equal justice," one need

64. PL, *supra* note 16, at 24; see also TJ, *supra* note 14, at 141 ("[T]o each according to his threat advantage is not a principle of justice.").

65. JF, *supra* note 41, at 16; see also Kymlicka, *supra* note 41, at 62 (explaining that the veil of ignorance "ensures that those who might be able to influence the selection process in their favour, due to their better position, are unable to do so").

66. PL, *supra* note 16, at 24; see also JF, *supra* note 41, at 18. That is, the parties "all have the same rights in the procedure for choosing principles; each can make proposals, submit reasons for their acceptance, and so on." TJ, *supra* note 14, at 19.

67. JF, *supra* note 41, at 18; see also PL, *supra* note 16, at 24 ("That the parties are symmetrically situated is required if they are to be seen as representatives of free and equal citizens who are to reach an agreement under conditions that are fair.").

68. JF, *supra* note 41, at 18.

69. *Id.* at 18.

70. *Id.* at 19.

71. *Id.* at 18-19.

72. See PL, *supra* note 16, at 79:

Citizens are equal in virtue of possessing, to the requisite minimum degree, the two moral powers and the other capacities that enable us to be normal and fully cooperating members of society. All who meet this condition have the same basic rights, liberties, and opportunities, and the same protections of the principles of justice.

not be particularly successful or well-spoken or remarkable in any way, nor must one's goals and ambitions be grand or meet with broad popular approval.⁷³ Ordinary people with modest aspirations, even unlikeable people with attributes and preferences others despise, are entitled to the same consideration and respect from society's social and political institutions as anybody else, as long as they are able to "understand, to apply, and to act from" fair terms of cooperation with others.⁷⁴

73. See TJ, *supra* note 14, at 19 ("[T]he parties in the original position are equal. . . . Systems of ends are not ranked in value . . .").

74. In his discussion of the nature of the good for persons, Rawls considers one "whose only pleasure is to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns." *Id.* at 432. Although we may find this preference bizarre, if it should prove that this person is perfectly sane and genuinely enjoys this activity for its own sake, we must concede that this is the definition of the good for this person. See *id.* at 432-33. And if this is so, presumably the grass-counter should be left to pursue his conception of the good without interference in just the same way as a violinist or a philanthropist should be left to pursue hers. This conclusion, it bears noting, in no way suggests we must value or encourage these various pursuits equally, but only that we must respect them equally as defining the good for free and equal individuals, and not condemn or seek to inhibit their pursuit simply because we may value them differently than does the one who pursues them. This is not, however, to say that in a liberal society no conceptions of the good would be excluded outright. To the contrary, those conceptions that are "in direct conflict with the principles of justice," such as "a conception of the good requiring the repression or degradation of certain persons on, say, racial or ethnic, or perfectionist grounds," may be actively discouraged by a liberal society. JF, *supra* note 41, at 154. Rawls here gives the examples of "slavery in ancient Athens or in the antebellum South." *Id.*; see also PL, *supra* note 16, at 187 ("[I]t is neither possible nor just to allow all conceptions of the good to be pursued (some involve the violations of basic rights and liberties)."). This inevitable exclusion of such conceptions of the good once the veil is lifted would be endorsed by the parties even knowing as they do that they themselves might turn out to hold a forbidden conception. For the parties would also know that they could turn out to be the targets of such a conception, and suffer at the hands of their fellows as a result. But still, it is not possible to say in advance which conceptions may be thus discouraged, for until the principles of justice are chosen, it is impossible to say which particular goals and preferences will conflict with them. For this reason, it would be premature to exclude any perspective from equal consideration in the original position on the grounds that the particular conception of the good it represents will prove to be inadmissible once the veil is lifted and the parties enter society as citizens. For discussion of the objection that the interests of potential criminals should be excluded from consideration in the original position on the ground that their conceptions of the good are morally contemptible, see *infra* note 166.

These constructs and assumptions model Rawls's commitment to what Samuel Scheffler has called "equality as a social and political ideal."⁷⁵ Together, they translate in Rawls's theory into a strict standard of impartiality, or what amounts to due consideration for the interests of all.⁷⁶ For if you are uncertain of what your social position might end up being once the veil is lifted, if all you know is that you have some social position and some "ideal of the good life" but not the particulars thereof,⁷⁷ you can only promote your own good if you also promote the good of all others. The conditions of the original position thus amount to a requirement that the parties choosing the principles of justice to govern the basic structure carefully consider the various options from the perspective of all possible social positions.⁷⁸ Because the parties do not know which social

75. Samuel Scheffler, *What is Egalitarianism?*, 31 *Phil. & Pub. Affairs* 5, 22 (2003). "As a moral ideal," Scheffler explains,

[equality] asserts that all people are of equal worth and that there are some claims that people are entitled to make on one another simply by virtue of their status as persons. As a social ideal, it holds that a human society must be conceived of as a cooperative arrangement among equals, each of whom enjoys the same social standing. As a political ideal, it highlights the claims that citizens are entitled to make on one another by virtue of their status as citizens, without any need for a moralized accounting of the details of their particular circumstances. Indeed, it insists on the very great importance of the right to be viewed simply as a citizen, and to have one's fundamental rights and privileges determined on that basis, without reference to one's talents, intelligence, wisdom, decision-making skill, temperament, social class, religious or ethnic affiliation, or ascribed identity.

Id.

76. See Brian M. Barry, *Justice as Impartiality* 11 (1995) ("Roughly speaking, behaving impartially . . . means not being motivated by private considerations.").

77. Kymlicka, *supra* note 41, at 64 (quoting Jeremy Waldron, *Theoretical Foundations of Liberalism*, 37 *Phil Q.* 127 (1987)).

78. Some critics have argued that the demands of the veil of ignorance neglect the constitutive role that the basic features of our identity—our gender, race, and native community, for example, not to mention our "special psychological propensities"—play in our capacity to deliberate. From this perspective, the notion of deliberating behind the veil makes no sense because behind the veil, parties will have been stripped of the very characteristics that make them who they fundamentally are, leaving no "selves" left to participate in the process of selecting principles of justice. See Seyla Benhabib, *The Generalized and the Concrete Other: The Kohlberg-Gilligan Controversy and Feminist Theory*, in *Feminism as Critique* 77, 89-90 (Seyla Benhabib & Drucilla Cornell eds., 1987);

position they will occupy once the veil is lifted, they would be led to consider what each set of proposed principles would have in store for them even if they ended up being situated among society's least powerful, least privileged, or most reviled members.

In this way, the framework of the original position and the principle of political legitimacy come together: we step behind the veil to be better able to consider the options from others' perspectives, free from our own inherent biases,⁷⁹ and once we have done so, we are collectively better equipped to justify the state's exercise of coercive power with reasons that "all citizens can publicly endorse in the light of their own reason."⁸⁰ In this sense, the design of the original position may be understood to accord equal consideration and respect to all members of society,

Michael J. Sandel, *Liberalism and the Limits of Justice* 27-28 (1982).

As Kymlicka explains, however, this objection mistakes the veil of ignorance for "a theory of personal identity." Kymlicka, *supra* note 41, at 62. The requirement that knowledge of our personal characteristics and social position be screened out by a veil of ignorance is not intended to model the way we would in fact deliberate in the real world. It is instead a heuristic device, a way to gauge whether proposed principles of justice really are impartial, whether they really give equal consideration to the interests of all affected parties. See *id.* (describing the original position as "an intuitive test of fairness, in the same way that we try to ensure a fair division of cake by making sure that the person who cuts it does not know which piece she will get"). It is not that there would ever be a conference of such artificial persons deliberating behind the veil. It is rather that any proposed principles, if they are to be affirmed as legitimate, must be shown by their proponents to be such that they *would* pass muster at such a conference, where such principles must be shown to be acceptable to anyone, whatever his or her social position.

In an effort to eliminate confusion on this point, Rawls in his later work emphasizes that the original position is "simply a device of representation," in which participants are to be seen as "artificial persons we fashion to inhabit the original position," representatives or trustees of free and equal citizens (and not the citizens themselves) who are to reach an agreement "subject to conditions that appropriately limit what they can put forward as good reasons." PL, *supra* note 16, at 25, 75.

79. In practice, of course, we can never divest ourselves of the knowledge of our own social position. The thought experiment of the original position simply challenges us to assess carefully the alternative principles from the perspective of individuals in society's least privileged positions, and to consider which, if any, of these principles are such that individuals in such positions might reasonably be expected to endorse them.

80. JF, *supra* note 41, at 91.

whatever their station.⁸¹ And, for this reason, the conclusions it yields regarding the deployment of coercive state power may be thought to be legitimate, whether that deployment is directed toward the (re)distribution of social goods or the punishment of criminal offenders.

C. Attributes and Actions: Morally Arbitrary Characteristics and the Distribution of Goods

Rawls's evident concern with eliminating the undue effect of arbitrary contingencies on the content of the principles of justice has generated some confusion as to the precise implications of this concern for the theory as a whole. In particular, it is not the case, as has sometimes been thought,⁸² that Rawls understands the central aim of a theory of justice to be the distribution of goods in a way that mitigates the effects of arbitrary contingencies on the

81. In his early review of *A Theory of Justice*, Ronald Dworkin read Rawls's theory to represent the view that "individuals have a right to equal concern and respect in the design and administration of the political institutions that govern them." Ronald Dworkin, *Justice and Rights*, 40 U. Chi. L. Rev. 500 (1973), reprinted in R.M. Dworkin, *Taking Rights Seriously* 150, 180 (1977). Although Rawls subsequently disclaimed this suggestion, it was the notion of justice as fairness "as a right-based view," *Justice as Fairness: Political not Metaphysical* (1985), reprinted in John Rawls: *Collected Papers*, supra note 40, at 388, 400 n.19, and not an ultimate commitment to the liberal ideal of equal concern and respect for all citizens that Rawls resisted. To the contrary, Rawls's response to Dworkin demonstrates that, for Rawls, the design and operation of the original position *does* add up to an affirmation of this ideal:

I think of justice as fairness as working up into idealized conceptions certain fundamental intuitive ideas such as [that] of the person as free and equal. . . . [T]hese fundamental intuitive ideas reflect ideals implicit or latent in the public culture of a democratic society. In this context, the original position is a device of representation that models the force, not of the natural right of equal concern and respect, but of the essential elements of these fundamental intuitive ideas as identified by the reasons for principles of justice that we accept on due reflection. As such a device, it serves first to combine and then to focus the resultant force of all these reasons in selecting the most appropriate principles of justice for a democratic society. (In doing this the force of the natural right of equal concern and respect will be covered in other ways.)

Id. at 400-01 n.19.

82. See, e.g., Kymlicka, supra note 41, at 70-71 (discussed in Scheffler, supra note 75, at 8-9).

ultimate distribution of goods.⁸³ There are, to be sure, a number of theorists who construe the aim of distributive justice in this way. These theorists, whom Elizabeth Anderson labels "luck egalitarians,"⁸⁴ "den[y] that a person's natural talent, creativity, intelligence, innovative skill or entrepreneurial ability can be the basis for legitimate expectations."⁸⁵ Instead, they take as their "core idea" that "inequalities in the advantages that people enjoy are acceptable if they derive from the choices that people have voluntarily made, but that inequalities deriving from unchosen features of people's circumstances are unjust."⁸⁶

This perspective, however, is not shared by Rawls,⁸⁷ as is clear from the content of the difference principle, the second principle of distributive justice he ultimately endorses.⁸⁸ On this principle, inequalities in the distribution of the primary goods are allowed so long as they benefit the least well-off member(s) of society. Consistent with this limitation, law-abiding individuals would thus be free to garner greater quantities of the primary goods—or squander the bundle they already have—and it is irrelevant to the justice of the result whether their capacities or tendencies for doing so are a product of voluntary choice or unchosen contingencies.

Consider, for example, the willingness to work hard. At various points, Rawls emphasizes his view that whether one possesses such a willingness is wholly the result of "social contingencies and natural fortune."⁸⁹ As he puts it, "[e]ven the willingness to make an effort, to try, and so to be deserving in the ordinary sense is itself dependent on

83. Scheffler identifies and challenges this misconstrual of Rawls's theory in his "What is Egalitarianism?" See Scheffler, *supra* note 75.

84. Elizabeth S. Anderson, *What Is the Point of Equality?*, 109 *Ethics* 287, 290 (1999).

85. Scheffler, *supra* note 75, at 6.

86. *Id.* at 5.

87. See *id.* at 25 ("[T]he underlying motivation for Rawls's theory of justice is not the general elimination of the influence of brute luck on distribution. . . . [He instead] aims to identify the most reasonable conception of justice to regulate the basic structure of a modern democratic society.").

88. For a full statement of Rawls's two principles of justice, see *supra* note 48.

89. TJ, *supra* note 14, at 73.

happy family and social circumstances.”⁹⁰ Yet once the principles of justice are selected and the veil is lifted, nothing in Rawls’s theory requires that the distribution of goods remain unaffected by an individual’s drawing on her willingness to work hard to increase her bundle of primary goods. To the contrary, so long as the inequalities generated in this way improve the welfare of the least well-off, the industrious members of society whose work has led to this improvement are entitled on the terms of justice as fairness to the concomitant increase in their own welfare brought about by their own hard work. And this is so, on Rawls’s theory, notwithstanding that an individual’s capacity for hard work is indisputably a function of morally arbitrary contingency.

Why would Rawls allow what he views as a patently arbitrary accident of fortune to influence the size of the bundles distributed according to the principles of justice? One possible answer lies in Rawls’s desire to create conditions in which the talents of naturally well-favored members of society could be harnessed for the greater good.⁹¹ He thus seeks to encourage such individuals to put their talents to work with the promise of greater bundles of goods than they would otherwise enjoy. But this explanation does not fully resolve the puzzle. For Rawls not only seems to accept the possibility that the possession and exercise of desirable yet arbitrary qualities could, consistent with the two principles, accrue to the *benefit* of the possessor, but he is also apparently willing to contemplate the possibility that the possession and exercise of undesirable arbitrary qualities could yield a *disbenefit* for the possessor without offending the demands of justice. Rawls thus appears untroubled by the example of the person with expensive tastes, who squanders her bundle of goods on “expensive wines and exotic dishes” and is left

90. Id. at 74.

91. “The premiums earned by scarce natural talents [through the application of the difference principle], for example, are to cover the costs of training and to encourage the efforts of learning, as well as to direct ability to where it best furthers the common interest.” Id. at 311.

with minimal resources,⁹² despite the fact that such expensive tastes may well arise from being “brought up in a wealthy family”⁹³—an experience that is surely an arbitrary accident of fortune.

The answer to this puzzle lies in a distinction which, although not explicitly articulated in Rawls’s theory, nonetheless lies at its core: that between the parties who possess (or lack) particular capacities, proclivities, and preferences, and citizens who act on the capacities, proclivities, and preferences they possess. Rawls is certainly committed to the view that whether or not one possesses particular characteristics is morally arbitrary.⁹⁴ For this reason, personal knowledge of one’s own characteristics is not permitted to influence the content of the principles of justice according to which society’s goods are to be distributed. Yet once the principles are chosen and the parties enter society as citizens, Rawls assumes that citizens have the capacity to conform their behavior to comport with their legitimate expectations.⁹⁵ He is thus quite willing to hold these citizens responsible for the consequences of acting on their preferences in a way that leaves them with inadequate resources,⁹⁶ notwithstanding

92. John Rawls, *Social Unity and the Primary Goods* (1982), in John Rawls: *Collected Papers*, *supra* note 40, at 369 [hereinafter SU].

93. Indeed, although he recognizes full well the accidental and thus morally arbitrary nature of our family circumstances, Rawls states quite plainly that “[w]e don’t say that because the preferences arose from upbringing and not from choice that society owes us compensation. Rather, it is a normal part of being human to cope with the preferences our upbringing leaves us with.” PL, *supra* note 16, at 185 n.15.

94. See TJ, *supra* note 14, at 311-12.

95. See SU, *supra* note 92, at 369.

96. Nowhere does Rawls himself draw such an explicit contrast between attributes and actions. Instead, he frames the idea in terms of citizens being “held responsible for their ends.” Id. (“[I]t is public knowledge that the principles of justice view citizens as responsible for their ends.”). But this difference in terminology aside, it is necessarily actions and not ends per se for which, on Rawls’s account, citizens are to be “held responsible.” For ascribing responsibility suggests holding to account, and it is not clear how, consistent with Rawls’s broader theory, citizens may be held responsible in this sense for their preferences unless and until they have acted upon them. Rawls certainly does not suggest, for example, that the interests of those people with expensive tastes, who prefer “expensive wines and exotic dishes” to “a diet of milk, bread, and

his recognition of the morally arbitrary circumstances from which these preferences would have arisen.

For Rawls, the assumption that individuals are properly held responsible for their actions flows directly from citizens' status as free and equal moral agents, possessed of the "moral power to form, to revise, and rationally to pursue a conception of the good."⁹⁷ Citizens are not merely "passive carriers of desires," but rather, "as moral persons[,] have some part in cultivating their final ends and preferences."⁹⁸ Moreover, holding individuals responsible in this way is a necessary feature of the well-ordered society. For were the principles of justice to compensate those who ultimately act on, say, their laziness or profligacy, such principles could render insecure the goods of those citizens who are not lazy or immoderate or who have not indulged those tendencies. For, as Rawls puts it, "[i]n any particular situation, . . . those with less expensive tastes have presumably adjusted their likes and dislikes over the course of their lives to the income and wealth they could reasonably expect," and it would be "unfair that [those citizens] now should have less in order to spare others from the consequences of their lack of foresight or self-discipline."⁹⁹

beans," *id.*, should be excluded from consideration in the original position on account of these preferences. To the contrary, the parties, who know the general facts about human society, will surely know that some citizens will turn out to have fancier tastes than others, and would thus, in selecting the principles, consider the possibility that they themselves will prove to have such tastes when the veil is lifted. Nor, once the veil is lifted, are citizens in any way held to account for preferences and ends to which they might remain committed but which, because of their circumstances, they are unable to realize without incurring great cost to themselves or imposing great harm on others and which they thus do not pursue. To the contrary, of such people, Rawls is likely to say that they are acting in a laudable and morally mature fashion, for by not externalizing the costs of their own bad choices, such individuals are behaving respectfully toward their fellows. On Rawls's view, it is simply the ones who fail to rein in their preferences, despite the fact that, in doing so, they create negative externalities for others, who must be held responsible and endure any negative consequences for themselves that would thereby arise.

97. *Id.*

98. See *supra* note 93. See also SU, *supra* note 92, at 369-70 ("[W]e must assume that citizens can regulate and revise their ends and preferences in light of their expectations of primary goods. This assumption is implicit in the powers we attribute to citizens in regarding them as moral persons.").

99. SU, *supra* note 92, at 370. As Scheffler explains:

Rawls is thus no luck egalitarian.¹⁰⁰ Luck egalitarians seek to compensate for the effects of morally arbitrary unchosen circumstances on the size of a person's bundle through the (re)distribution of society's goods. For Rawls, in contrast, individuals are responsible for the effects of their own choices, even if those choices derive from circumstances that are themselves morally arbitrary. To ensure that the principles governing the distribution of goods are unaffected by morally arbitrary contingencies but sensitive to individuals' choices and actions, Rawls divides the process of *selecting* the principles from the process of *applying* them in concrete cases, and sees to it that it is not the ultimate distribution but the terms of the original position—and these terms only—that are manipulated to nullify the effects of morally arbitrary contingencies. “The arbitrariness of the world,” Rawls maintains, “must be corrected for by adjusting the circumstances of the initial contractual situation.”¹⁰¹ However, once the veil is lifted and the parties enter society as citizens, it is their actions and not their mere attributes that in a just society will determine any change in the size of their bundles. Thus, any benefits or burdens that subsequently accrue to individuals in accord with the two principles of justice are traceable not to morally arbitrary contingencies, but to actions for which individuals are properly held responsible.¹⁰²

People are asked to accept responsibility for their ends, in Rawls's sense, not because the metaphysics of the will makes it fitting that people should bear the costs of their choices, but rather because it is reasonable to expect people to make do with their fair shares. And what makes shares fair, according to Rawls, is not that they compensate people for all unchosen disadvantages while leaving them to bear the costs (or reap the rewards) of their voluntary choices. Shares are fair when they are part of a distributive scheme that makes it possible for free and equal citizens to pursue their diverse conceptions of the good within a framework that embodies an ideal of reciprocity and mutual respect.

Scheffler, *supra* note 75, at 27-28.

100. See *id.*

101. TJ, *supra* note 14, at 141.

102. True, the actions one is able to undertake while pursuing one's conception of the good will to a great extent be influenced by the nature of one's attributes, which are themselves, Rawls recognizes, the product of morally arbitrary contingencies. See *supra* note 24. But just as, for Rawls, “it is a normal part of

D. Uncertainty and Risk: Maximizing Personal Welfare from Behind the Veil

This, then, is the construct within which the parties are, on Rawls's framework, to select the principles of justice. But before we are to be able to apply this construct to the problem of punishment, we need to answer one final, essential question: how do the parties, deliberating in the original position behind the veil of ignorance, conduct their deliberations? Their task is to select, from among various proposed options, the principles that are to guide society's basic structure. Yet how are they to evaluate these options? And on what basis are they to choose among them?

It is here that Rawls's idea of the primary goods becomes centrally relevant. Rawls defines the primary goods as those "various social conditions and all-purpose means" that citizens need in order to develop the moral powers they need to pursue their own conceptions of the good in a shared world.¹⁰³ Although the parties do not know their own particular conceptions of the good, they do know that they have some such conceptions which they will want to pursue.¹⁰⁴ They thus know that it will go better for them in society if they can get for themselves as large a bundle as possible of the primary goods. The parties will therefore select those principles that will ensure them, once the veil is lifted, as large a bundle as possible of the available primary goods. The question then becomes: how are the parties to make this determination?

being human to cope with the preferences our upbringing leaves us with," PL, supra note 16, at 185 n.15, so too is it a part of our moral obligations to others in society to figure out how best to use our bundles of goods to realize our ends while ensuring that our actions do not unfairly burden others. As long as we are in possession of the two moral powers, Rawls assumes that we have the capacity to shape our conduct to meet this obligation and must bear the consequences should we fail to do so.

103. JF, supra note 41, at 57.

104. See TJ, supra note 14, at 142 ("[T]he parties do not know their conception of the good. This means that while they know that they have some rational plan of life, they do not know the details of this plan, the particular ends and interests which it is calculated to promote.").

1. Gambling under Uncertainty

One might think that at least some parties, deliberating rationally behind the veil, might approach this problem by anticipating that in any society, there are likely to be more powerful people and less powerful people, and that the more powerful one is, the more likely one is to enjoy more of what the parties seek. Because the benefits of being among the most powerful might be expected to be considerable, it may seem reasonable to imagine that some of the parties—at least the gamblers among them—might decide to take the chance that they themselves would wind up among this most favored group. Doing so would lead this group of risk-takers to opt for principles that greatly enhance the position of the best off, even at the expense of other, less fortunate citizens, whom they gamble they will not turn out to be.

This risk-taking posture, however, is not one that rational agents, seeking to maximize their future prospects from behind the veil, would adopt. The reason is not, as might be thought, because the parties are unusually “risk-averse”;¹⁰⁵ to the contrary, under the conditions of the original position, even those with a general disposition to take risks would forbear doing so here. For if one is to take a gamble in this way—particularly a gamble on which rides so much of importance¹⁰⁶—one needs some way to gauge the probabilities one faces. If we knew, for example, that were we to choose a certain principle, society would prove once the veil is lifted to have one rich person for every poor person, a gambling type might be inclined to embrace that

105. To the extent that one's disposition toward risk is a feature of one's general character and personal conception of the good, we can expect attitudes to risk to vary among the parties to the same degree as they vary among the general population. See S.L. Hurley, *Natural Reasons: Personality and Polity* 374 (1989). But because the particular nature of each party's personal attributes is hidden behind the veil, no participant has any way of knowing his or her particular disposition toward risk, and thus none can allow this disposition to govern his or her approach to the selection of the principles.

106. See *id.* at 381 (describing the parties' choice of the principles of justice as “a one-off decision in which [their] entire life is at stake”).

principle, and take the chance that she will end up a rich person, even if this meant an equal chance she could end up poor. The problem, however, is that the parties behind the veil have access to no such insights as to the eventual make-up of society on the various alternative principles; the veil not only keeps them ignorant of their own personal attributes, but it also keeps from them any knowledge of "the particular circumstances of their own society. . . . its economic or political situation, or the level of civilization and culture it has been able to achieve."¹⁰⁷

The parties thus have no basis at all on which to determine their chances of winding up well-favored or ill-favored in their share of the primary goods. Instead, on this score, they face conditions of total uncertainty. And, as Susan Hurley explains, far from tempting actors to roll the dice, conditions of total uncertainty consistently lead even risk-takers to avoid taking chances. Indeed, she describes such "uncertainty aversion" as so common that it "may reasonably feature in a theory of human nature."¹⁰⁸ Given such an aversion, under the conditions of uncertainty the parties face,¹⁰⁹ we can thus expect the parties not to take a gamble on an unpalatable result.¹¹⁰

107. TJ, *supra* note 14, at 137.

108. Hurley, *supra* note 105, at 376; see also *id.* at 374 (describing the "experimental results" demonstrating this preference as "remarkably robust"); *id.* at 374-77 (discussing experiments that yielded this insight as well as variations connecting it to the task facing Rawls's deliberators).

109. Under these circumstances, not only do the parties have no way of gauging their probable success, but they also have no way of knowing just how bad the worst result may be. And as Hurley argues, under these circumstances, how could any party "have reason to base its decisions on probabilistic considerations, which gain appeal from consideration of the long run of cases, when the only decision it is called upon to make is a one-off decision in which its entire life is at stake?" *Id.* at 381.

110. Other commentators have taken a different view of this question. Harsanyi in particular has argued that under the conditions of the original position, the parties would assume that all outcomes are equally possible. See John C. Harsanyi, *Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory*, 69 *Am. Pol. Sci. Rev.* 594, 598-600 (1975) (reviewing John Rawls, *A Theory of Justice* (1971)) (explaining and defending what he calls the "equiprobability assumption"). Rawls himself, however, has made clear that he considers uncertainty rather than risk to be the defining feature of the parties' situation. See JF, *supra* note 41, at 106 ("[W]e view the

How then are the parties to go about identifying the principles by which they can most comfortably live? Rawls argues that, given the conditions of the original position, the parties would be best served by reasoning according to the "maximin rule."¹¹¹ According to maximin, parties concerned to maximize their long-term prospects under conditions of uncertainty ought to proceed on the assumption that they could end up occupying the social position of the worst-off. They should then select from among the available options that which guarantees the best possible result for the occupants of that position,¹¹² and not be "much concerned for what might be gained" by those whose positions turn out to be more congenial.¹¹³ In short, to ensure adequate consideration of the worst possibilities a particular proposed principle has in store, each participant in the original position would be well-advised to follow Rawls's advice, and consider the implications of each principle as if, once the veil is lifted, "his enemy is to assign him his place [in the social order]."¹¹⁴

That uncertainty rather than risk is the appropriate description of the situation facing the parties is clear from

parties as faced with uncertainty rather than risk."); see also *id.* at 106 n.29 (citing Hurley as having "an instructive discussion of risk aversion and uncertainty and their relation to the maximin rule"); TJ, *supra* note 14, at 169 ("we are willing [to take great risks for ourselves] only when there is no way to avoid these uncertainties. . . . Since the parties have the alternative of the two principles of justice, they can in large part sidestep the uncertainties of the original position.").

111. TJ, *supra* note 14, at 152; see also Hurley, *supra* note 105, at 376-77 ("[w]hen there are no known risks relating to specified goods to set against aversion to uncertainty, as is the case in the original position, maximin behaviour may be justified as the closest one can come to avoiding uncertainty.").

112. See TJ, *supra* note 14, at 152-53 (explaining that maximin reasoning advises the parties to "rank alternatives by their worst possible outcomes" and "to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others").

113. JF, *supra* note 41, at 98. This approach will ensure that the parties have considered the full range of hardships to which they might be subject, should they turn out to occupy the worst-off positions in society. See Shklar, *supra* note 3, at 35 ("[T]he most reliable test for what cruelties are to be endured at any place and any time is to ask the likeliest victims, the least powerful persons, at any given moment and under controlled circumstances.").

114. TJ, *supra* note 14, at 152.

the purpose Rawls intends for his deliberative framework. As we have seen, the original position with its veil of ignorance is intended to model conditions of strict impartiality. No one is to know the circumstances of his or her own place in the world, because should they have access to this information, the parties might be led to select principles that would favor people like themselves instead of giving full consideration to the implications of the alternatives from all possible perspectives. Rawls does, it is true, harness for this purpose the intuitively appealing idea that self-interest should govern selection. He does so, however, not in order to affirm self-interest as the appropriate basis for momentous decisions of this kind, but because he perceives that combining a self-interested effort to maximize one's own welfare with the veil of ignorance "achieves the same purpose as benevolence."¹¹⁵ As Rawls puts it, "this combination of conditions forces each person in the original position to take the good of others into account. . . . The feeling that this conception of justice is egoistic is an illusion fostered by looking at but one of the elements of the original position."¹¹⁶ The condition of uncertainty the parties face behind the veil should thus be understood as an effect of this combination of stipulated conditions. The ultimate intention is not to ensure that the parties protect their own individual prospects, but that they consider whether the possible results of any alternative would be justifiable when viewed from any social position.¹¹⁷

2. The Strains of Commitment

The parties, under conditions of uncertainty, will thus reason according to maximin. By adopting this approach,

115. *Id.* at 148.

116. *Id.*

117. For "when the possibilities in question are all human lives, and all of them will in fact be lived," it should be of no relevance to the truly impartial that a particular experience will be "*mine* (rather than anyone else's)." Hurley, *supra* note 105, at 382 (emphasis in the original).

they are able to protect as far as possible their fundamental interest in guaranteeing themselves the largest possible bundle of primary goods. In addition to their concern to maximize their prospects with respect to these primary goods, however, the parties have a further fundamental interest—that in identifying a set of principles to which all members of society may willingly adhere over the long term. For unless the principles that are selected satisfy this condition, there can be no guarantee of social stability over time. Instead, the character of the just society promised by the principles of justice ultimately selected would be at constant risk of being undermined whenever anyone found herself unable to honor her initial commitment.¹¹⁸ Recognizing this interest in long-term social stability, Rawls argues that, in considering the available principles, the parties in the original position will be sure to attend to what he calls the “strains of commitment.”¹¹⁹ That is, the parties will be sure not to enter into agreements the potential consequences of which could cause too great a strain, and would thus be too heavy to bear, even for those who agreed in good faith to honor the agreement reached, however things turned out.¹²⁰

As Rawls explains it, the parties’ concern with the strains of commitment reflects the contractarian nature of the agreement reached in the original position. The parties in the original position do not take a vote on the principles to govern society, with, say, majority opinion winning the day. They are instead forging a contract, under which each party pledges to adhere to the terms of the agreement.¹²¹ This is,

118. See TJ, *supra* note 14, at 145 (“[T]he parties can rely on each other to understand and to act in accordance with whatever principles are finally agreed to. Once principles are acknowledged the parties can depend on one another to conform to them. In reaching an agreement, then, they know that their undertaking is not in vain . . .”).

119. See TJ, *supra* note 14, at 175-76; John Rawls, *A Reply to Alexander and Musgrave* (1974), reprinted in *John Rawls: Collected Papers*, *supra* note 40, at 232, 249-50 [hereinafter *Reply*]; JF, *supra* note 41, at 102-03.

120. See *Reply*, *supra* note 119, at 250.

121. *Id.* at 249 (“[R]eaching a unanimous agreement without a binding vote is not the same thing as everyone’s arriving at the same choice, or forming the same intention.”).

moreover, a contract with high stakes; the parties are "choosing once and for all the standards which are to govern [the] life prospects" of the citizens they represent.¹²² Because "the original agreement is final and made in perpetuity, there is no second chance."¹²³ The character of the agreement as a mutual undertaking to bind citizens in perpetuity gives the parties assurance that, once the veil is lifted, each can count on the others to adhere to the terms of the agreement over the long term. The parties thus cannot make this initial pledge unless "[they] intend to honor it," which in turn means that they must have "reason [to] believe that [they] can do so."¹²⁴ For this reason, "no one is permitted to agree to a principle if they have reason to doubt that they will be able to honor the consequences"¹²⁵

Given the extent of the coercive power at the disposition of the state, it may seem odd that the parties are concerned at all with avoiding agreements they could prove unable to honor. For once an agreement is reached as to the content of the principles and the veil is lifted, the state would surely have sufficient power to ensure the implementation of the agreement, whether or not citizens willingly endorse the implications for themselves personally of the principles ultimately selected. This way of viewing the matter, however, misconstrues the point of the strains of commitment. The parties' concern is not to ensure acquiescence in the terms of the original agreement in order to prevent grumbling among the citizenry or to make the distribution of goods run smoothly, without added hassle from uncooperative individuals. Their concern is instead with the quality of the agreement. That is, the aim is to identify principles to which all can willingly adhere over the long term. Otherwise, all that would remain is a powerful coercive state imposing a

122. TJ, *supra* note 14, at 176.

123. *Id.*

124. Reply, *supra* note 119, at 250. Rawls thus concludes that "[t]he class of things we can agree to is included within but smaller than the class of things that we can rationally choose." JF, *supra* note 41, at 102.

125. Reply, *supra* note 119, at 250.

particular vision of society on a resentful and recalcitrant population—surely no one's vision of a just society. Under such unhappy circumstances, as Rawls suggests, citizens would not affirm the society as just, but would instead become at best “distant from political society,” and “withdrawn and cynical,” and at worst “sullen and resentful, . . . [ready] to take violent action in protest against [their] condition.”¹²⁶ Because the parties strongly wish to avoid either outcome, they seek principles the results of which all citizens can agree are just and fair and have no cause to resent.¹²⁷ The parties will thus be careful to attend to the strains of commitment when selecting the principles of justice.

These, then, are the features of Rawls's model that will be of the greatest relevance to our efforts to apply this model to the problem of punishment. But before we can put the model to work in this way, we must first get clear on the relevant range of social positions the parties would view as among those they could end up occupying once the veil is lifted. We thus turn first to determining this relevant range. In doing so, we shift our focus from Rawls's well-ordered society to a society marked by partial compliance, for only in the context of a partially compliant society does the question of what constitutes just punishment become at all relevant. Specifically, in what follows, I consider three possible social positions in a society marked by crime and (possibly)¹²⁸ punishment: crime victim, wrongfully convicted innocent, and guilty offender. Because the parties' willingness to consider the consequences of possible principles for any social position turns on whether they imagine that they themselves could end up occupying that position once the veil is lifted, the extent of the parties' uncertainty with respect to these particular positions will profoundly affect the content of the final agreement.

126. JF, *supra* note 41, at 128.

127. See Reply, *supra* note 119, at 250.

128. I say “possibly” here because whether and to what extent punishment should be imposed is precisely the question the parties must answer.

III. PERSPECTIVES ON PUNISHMENT IN A PARTIALLY COMPLIANT SOCIETY

A. *From Ideal Theory to Partial Compliance*

The question now becomes how to apply this Rawlsian framework to the task of identifying principles of just punishment. First, as we have seen, we must dispense with Rawls's idealized picture of a well-ordered society. This move, although perhaps unanticipated by Rawls himself, is necessary for our purposes. For if we are to arrive at meaningful principles, we must consider the question from the perspective of parties who have a stake in ensuring their own self-protection, whether against criminal offenders or against the state, and citizens of a well-ordered society face little if any threat from these quarters. Indeed, in a well-ordered society, the precise contours of the punishments prescribed for various infractions would matter little to the parties, for under conditions of strict compliance, as Rawls puts it, "[t]hese mechanisms will seldom be invoked"¹²⁹ To render the Rawlsian framework meaningful to the task at hand, we must therefore exchange at least some features of ideal theory for those of partial compliance.

For our purposes, three features of institutional and public life in a Rawlsian well-ordered society are particularly relevant:

129. TJ, *supra* note 14, at 577. Because citizens in a well-ordered society may be expected to act justly, honoring the "basic natural duties, those which forbid us to injure other persons in their life and limb, or to deprive them of their liberty and property," *id.* at 314, crime—at least crime involving violations of this sort—will happen rarely if ever. If a scheme of state sanctions is to serve any purpose in such a world, it would be to assure citizens that, like themselves, others are playing by the rules, for absent such assurance, citizens "may suspect that some are not doing their part, and so they may be tempted not to do theirs." *Id.* at 240. Punishment thus exists in such a well-ordered society less as a manifest invocation of state power than as an idea, albeit one serving a valuable stabilizing function. For discussion on this point, see *id.*

1. citizens act justly and do their part in upholding just institutions, strictly complying with the demands of justice and honoring their duties and obligations toward fellow citizens;¹³⁰
2. the rule of law obtains, the criminal justice system operates efficiently and effectively, and the laws of the land are administered by state agents consistently and impartially; and¹³¹
3. the principles and institutions ordering the distribution of society's basic primary goods,¹³² which we can think of as society's "background conditions," are just and known to be so.¹³³

In what follows, I consider from the perspective of the parties in the original position the implications of altering these three conditions of a well-ordered society while holding constant all other relevant features of such a society. The resulting social context, to which I will refer as the "partially compliant society," is like Rawls's well-ordered society in all respects, save the three just enumerated. Instead, in a partially compliant society, not all citizens may be relied on to act justly toward others; the institutions and operations of the criminal justice system are flawed and untrustworthy, and are recognized as such; and society's background conditions are unjust and known to be so.

As we will see, under these altered conditions, the parties would no longer be indifferent to the nature and extent of the exercise of the state's power to punish criminal offenders. The precise character of the parties' attitudes toward state punishment, however, will depend on how they anticipate being personally affected. Once they enter society as citizens, might they themselves wind up as victims of crime? Or as targets of state punishment? It is important that we determine how the parties would

130. See *id.* at 8-9, 314-15.

131. See *id.* at 235.

132. See *supra* note 41.

133. TJ, *supra* note 14, at 453-54.

answer this question, for the range of positions the parties would think they could occupy once the veil is lifted will greatly affect the principles of punishment they would ultimately endorse.

In this part, therefore, I consider the implications for the parties' perspectives on state punishment of altering the fortunate conditions of the well-ordered society along the three dimensions listed above.¹³⁴ The first step in so doing is to determine the parties' priorities when it is a partially compliant society and not a well-ordered one that they expect to enter.

B. The Priority of Security and Integrity

For purposes of their deliberations behind the veil, we assume that the parties are concerned exclusively with their own personal welfare. As we have seen, they are able to manifest this concern despite deliberating in the dark

134. It might be thought that polling data, rather than a theoretical treatment of this issue, would be the best way to get at what we seek here. Using polling data, that is, would show us what real people actually think about crime and punishment under realistic social conditions. The problem with this approach for our purposes is that we seek a sense of what people might think when viewing the issues impartially, as if they didn't know whether others or they themselves occupy relevant social positions. Survey respondents respond as themselves to questions asked; they make no attempt also to consider the questions from all other relevant perspectives. If we seek impartiality, therefore, this approach would not serve.

There may, of course, be some groups which, if surveyed, would give us results close to those we seek; perhaps, for example, responses to surveys on crime and punishment would be most fruitful for our purposes were they to come from, say, poor minority communities, whose residents are more likely than other members of society to be both victims of crime and targets of punishment. See Callie Marie Rennison, U.S. Dep't of Justice, *Criminal Victimization 2001* 3 (2002), <http://www.ojp.usdoj.gov/bjs/pub/pdf/cv01.pdf>; see also *supra* p. 311 and notes 8-9 (noting the disproportionate targeting for punishment of indigent people of color). Notice, though, that even here, the data's relevance for our purposes would depend upon their consistency with the standards of impartiality modeled by the veil of ignorance. This fact indicates the prior importance of the theoretical framework I develop. As for the approach itself, to seek out actual people whose particular social positions render them capable, when responding as themselves, to deliberate in something like the way the veil of ignorance requires is perhaps one way to frame the search for legitimate principles of punishment. In this paper, however, I adopt another.

about the details of their own personal identity, social position, affective ties to particular others, and conception of the good. For although they do not know the content of their own particulars, the parties do know that they have *some* social position, *some* personal identity, *some* affective ties to (as yet unidentified) particular others, and *some* conception of the good life that it will be important for them to pursue. The parties will therefore seek those principles that will foster the best climate for each person to realize a meaningful life, however he or she comes to understand that notion once the veil is lifted.

To do so, of course, the parties first must establish their priorities, the interests and goals that would be most valuable to them as citizens in society. Addressing this question in the distributional context, Rawls considers two candidates for the parties' highest priority: liberty and material resources.¹³⁵ Rawls's discussion of these candidates indicates that, although generally assuming that the parties' greatest interest is in ensuring the broadest possible scope for the unhindered pursuit of their own conception of the good, he also recognizes that individuals require a certain level of material resources if such pursuit is to be possible. Rawls argues that under unfavorable social conditions, in which some portion of the citizenry has insufficient material resources to meet their basic needs, the parties might well be willing to trade off some measure of liberty in order to increase their material well-being. It is only as "the marginal significance for [their] good of further economic and social advantages diminishes relative to the interests of liberty" that the parties would come to give priority to the interest of liberty, that good which would allow them to define and shape their own lives in the ways they choose.¹³⁶

135. See TJ, *supra* note 14, at 542-43.

136. *Id.* at 542; see also *id.* at 543 ("Until the basic wants of individuals can be fulfilled, the relative urgency of their interest in liberty cannot be firmly decided in advance. . . . But under favorable circumstances the fundamental interest in determining our plan of life eventually assumes a prior place.").

Yet where the goal is not to establish principles of distributive justice but instead to establish principles of just punishment, the parties would not shift their priorities depending on society's level of material well-being.¹³⁷ Here too we assume that the parties' highest interest is in ensuring the greatest possible scope for the unfettered pursuit of their aims. But in a social context in which both crime and punishment exist,¹³⁸ the greatest obstacle to such liberty is not a lack of material resources but the compromising by others of the subset of liberties to which I will refer collectively throughout this article as "security and integrity": security from assault on and interference with one's physical and psychological integrity and well-being. These goods, which as Rawls suggests are among the "basic liberties" that are "necessary if the other basic liberties are to be properly guaranteed,"¹³⁹ are necessarily prior to the possibility of any pursuit of one's goals. For without the protection of one's security and integrity in this sense, even the provision of adequate material resources

137. True, the state's ability to realize the demands of the principles in practice will depend on the level of societal resources. But this is a subsequent consideration, not taken into account by the parties in the original position, who, after all, know nothing of the details of their own society's level of development when selecting the principles.

138. For purposes of this part, I am assuming that the parties would endorse some form of state punishment when selecting principles of punishment for a partially compliant society. In part IV below, I consider the possibility that the parties might adopt a no-punishment principle, and ultimately argue that they would not.

139. JF, *supra* note 41, at 113 (placing the "basic liberties" of "liberty and integrity (physical and psychological) of the person" among those that are "necessary if the other basic liberties are to be properly guaranteed"). There are other basic liberties that would arguably be necessary to ensure the broadest possible scope for the pursuit of the parties' particular conceptions of the good. However, this particular liberty, which Rawls labels "the liberty and integrity of the person," is the primary liberty that is threatened by criminally-minded fellow citizens and which thus may be answered by a system of criminal punishment. The other liberties Rawls emphasizes, including "freedom of thought and liberty of conscience[, and] the political liberties and freedom of association," PL, *supra* note 16, at 291, are all liberties the parties must claim against the government and secure through the shaping of the political institutions of the basic structure. Consideration of these other liberties, however, is beyond the scope of this article.

can mean little.¹⁴⁰ For this reason, we can expect the parties' highest priority when contemplating the content of the principles of punishment to be that of self-protection, and in particular the protection of their security and integrity.

Because security and integrity are the necessary preconditions for the exercise of all other basic liberties, and because the conception of the good of one who is deprived of these goods will inevitably be compromised—and, depending on the degree of deprivations, perhaps seriously compromised—the parties would therefore seek those principles that promise once the veil is lifted to maximize their position measured in terms of these particular goods. And because different principles for governing the basic structure will have different implications for citizens' security and integrity, the parties will be particularly attentive to the nature of these differences in their deliberations. Ultimately, they can be expected to seek the principles of punishment they believe will best protect these goods, and to avoid those principles that put the citizens they represent in a less desirable position vis-à-vis these goods than the available alternatives.

The place of security and integrity among the preconditions for the maintenance of an adequate scheme of liberties means that the principles of punishment the parties select for a partially compliant society would form a part of Rawls's first principle of justice. True, Rawls's first principle, which accords to all members of society a "fully adequate scheme of equal basic liberties" where such a scheme is compatible with the "same scheme of liberties for all,"¹⁴¹ assumes a well-ordered society. Yet even a non-ideal society marked by conditions of unequal liberty¹⁴² may

140. There may of course be extreme exceptions, for example, one who seeks incarceration to receive medical care otherwise unavailable to the uninsured. I acknowledge that such extreme exceptions might exist, but in general I believe the main point to be sound.

141. JF, *supra* note 41, at 42.

142. Under conditions of partial compliance, there will necessarily be unequal

strive for justice. At the very least, such a non-ideal society may attempt to ensure the broadest possible protection for the most urgent and fundamental of citizens' interests. Thus, even recognizing the non-ideal nature of society, the parties seeking greater conditions of justice could—and would—make a priority of protecting their security and integrity. Given the priority of these most basic of primary goods, the parties seeking the principles of punishment for a partially compliant society would refuse to accept a diminution in their security and integrity in exchange for increased social and economic benefits.¹⁴³

C. Locating Crime Victims and Convicted Offenders Behind the Veil

Assuming the conditions of a partially compliant society, the parties' priority when considering which principles of punishment to select would thus be to maximize the protection of their security and integrity. How they go about doing so, however, will depend greatly on how they understand the nature of the threat to their share of these goods. Certainly, as the parties would know, crime, and especially violent crime, robs victims of the goods of security and integrity. It compromises their peace of mind, makes them fearful and tentative, and leaves them possibly physically violated and very likely psychologically traumatized even in cases where no physical harm was caused.¹⁴⁴ But so too would the parties,

liberties—and this would be so even were the parties to reject any principles authorizing the state to punish criminals. For even absent state punishment, in a partially compliant society there will inevitably be crime, and there will therefore be crime victims who suffer a diminution in their security and integrity—goods that number among the basic liberties—while others do not.

143. This priority is consistent with the "[p]riority [r]ule" Rawls attaches to his first principle of justice, that "the principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty." TJ, *supra* note 14, at 250.

144. Viewed from the perspective of the crime victim, punishment not only provides protection from physical and psychological harm, but as well provides the further benefit of affirming the victim's worth, and the wrongfulness of the criminal act done to them. This affirmation of crime victims reinforces their

with their access to the “general facts about human society,”¹⁴⁵ know that state punishment—and incarceration in particular¹⁴⁶—strips its targets of their security and integrity, keeping them in extended conditions of severe regimentation and physical vulnerability, conditions that can lead to physical violation and create a risk of ongoing psychological trauma and fear even if no physical harm is ever imposed.¹⁴⁷ The parties would thus need to consider whether they could end up occupying two possible social positions once the veil is lifted: crime victim, and target of state punishment. Either one of these positions would mean for its occupants a serious compromise of their security and integrity. We therefore ask: would the parties be uncertain as to whether they could end up, once the veil is lifted, as crime victims or convicted offenders facing punishment?

Certainly, once the veil is lifted, the parties would face uncertainty as to whether they could end up as crime victims. For if crime victims turn out to be chosen at random, this randomness would create such uncertainty. And even if, when the veil is lifted, victims of crime turn out to be concentrated in certain social groups, say, certain

security and integrity by assuring them that they and their well-being are valued by the collective. These considerations would necessarily be factored into calculations as to how particular punishments would impact the security and integrity of the relevant parties.

145. TJ, *supra* note 14, at 137.

146. For purposes of this article, I leave aside the question of capital punishment. It bears noting, however, that offenders sentenced to death also typically spend many years behind bars, thus suffering like compromises to their security and integrity even apart from the extreme ultimate deprivation execution represents.

147. Others have asserted an additional effect of punishment: the benefit to the offender of affirming his or her moral autonomy and status “as an individual worthy of respect,” Morris, *supra* note 35, at 157; see also Morris, *supra* note 21, at 74-93 (arguing that guilty offenders have a right to be punished and that “the denial of this right implies the denial of all moral rights and duties”), and of allowing the offender the opportunity to atone for and thereby expiate his wrongful acts. See Stephen P. Garvey, *Punishment as Atonement*, 46 UCLA L. Rev. 1801 (1999). Although I acknowledge the arguments in favor of such a view, I nonetheless assume that, from the perspective of the parties behind the veil, this view of punishment’s advantages to the offender would make it no more attractive to the parties than it would otherwise be.

racess or classes, the veil at this point prevents the parties from knowing not only which social groups those will be, but also to which social groups they themselves will belong. The parties thus lack any basis for estimating probabilities as to their own likelihood of ending up crime victims and for this reason would face uncertainty on this question. And, given this uncertainty, the parties would assume that they could end up occupying the position of crime victim. They would therefore seek principles that would maximize the position of crime victims in terms of their goods of security and integrity.

What, however, of the position of target of state punishment? Would the parties deliberating behind the veil face uncertainty as to whether they could find themselves, once the veil is lifted, convicted offenders facing punishment at the hands of the state? It is my contention that they *would* face such uncertainty. To make this case, however, it is necessary to respond to two different arguments that might be made for the opposite conclusion. In what follows, I address each in turn.

1. Criminal Offenders and the Moral Powers

Although the parties are ignorant of their own attributes and personal particulars, there are certain features of their own nature as moral beings that the parties do know. Most significantly for our purposes, the parties know that they all possess to the "requisite minimum degree," the capacity for a sense of justice and the capacity for a conception of the good.¹⁴⁸ That this is so might lead one to conclude that, because criminal offenders must by definition lack the capacity for a sense of justice—that is, "the capacity to understand, to apply, and to act from . . . the fair terms of social cooperation"¹⁴⁹—the parties

148. It is the possession of these basic capacities to the "requisite minimum degree" that for Rawls signifies the status of the possessor as a citizen entitled to "be counted as a full and equal member of society in questions of political justice." PL, *supra* note 16, at 302.

149. JF, *supra* note 41, at 18-19.

could be certain from the outset that they would not wind up convicted offenders facing state punishment once the veil is lifted.

This conclusion, however, would be mistaken. For to read this stipulation of a baseline possession of the two moral powers to exclude all criminal offenders from the original position would be to make the requirement of this capacity far more exacting than Rawls himself intends. Rawls in fact sets a deliberately low threshold in terms of the possession of the two moral powers, to which he refers collectively as "moral personality."¹⁵⁰ He "stresse[s] that the sufficient condition for equal justice, the capacity for moral personality, is not at all stringent," and that "[t]here is no race or recognized group of human beings that lacks this attribute."¹⁵¹ Although he concedes that "individuals presumably have varying capacities for a sense of justice," Rawls makes clear that "this fact is not a reason for depriving those with a lesser capacity of the full protection of justice."¹⁵²

The notion that one may have a minimal capacity for a sense of justice yet still meet the threshold requirements of the original position may seem odd. Surely, one is either capable of adhering to the fair terms of social cooperation with others, or one is not. There is, however, a difference between, on the one hand, always in fact acting out of a sense of justice according to fair terms of social cooperation, and, on the other hand, being a moral agent who sometimes fails to do so, yet who nevertheless understands the importance of so acting and who possesses the potential to act in these ways were the circumstances more congenial. And it is only the latter—the possession of the *capacity* for a sense of justice along with the *capacity* for a conception of the good—that determines one's admissibility to the original position.

Indeed, that it is the potential and not the full realization of one's moral powers that is required for

150. TJ, *supra* note 14, at 505.

151. *Id.* at 506.

152. *Id.*

participation in the original position is the only plausible understanding of Rawls's stipulation as to the basic moral minimum required of the parties; were it otherwise, no one would qualify. For as Rawls makes clear, even the citizens of a well-ordered society are no angels. To the contrary, they are inclined toward self-interest.¹⁵³ They act contrary to the "principles of right and justice," perhaps through cheating or cowardice, occasioning (appropriate) feelings of guilt and even shame.¹⁵⁴ Under the right circumstances, they can be expected to evade rules necessary to maintain collective order.¹⁵⁵ And, at times, even they violate the fair terms of social cooperation; as Rawls acknowledges, in a well-ordered society, even strict compliance is not perfect compliance.¹⁵⁶ Indeed, while their moral powers may lie above the requisite minimum level, citizens of the well-ordered society are always in need of further development in this regard.¹⁵⁷ Yet despite their limitations, Rawls considers these flawed individuals qualified to participate

153. See *id.* at 5 ("[M]en's inclination to self-interest makes their vigilance against one another necessary . . .").

154. *Id.* at 446.

155. See *id.* at 240:

[E]ven in a well-ordered society the coercive powers of government are to some degree necessary for the stability of social cooperation. For although men know that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. . . . The suspicion that others are not honoring their duties and obligations is increased by the fact that, in the absence of the authoritative interpretation and enforcement of the rules, it is particularly easy to find excuses for breaking them. Thus even under reasonably ideal conditions, it is hard to imagine, for example, a successful income tax scheme on a voluntary basis.

156. See JF, *supra* note 41, at 13 (explaining that "[s]trict compliance means that (nearly) everyone strictly complies with, and so abides by, the principles of justice," thus stipulating "realistic, though reasonably favorable, conditions") (parenthetical in the original).

157. According to Rawls, "the aim of the parties is to agree on principles of justice that enable the citizens they represent to become full persons, that is, adequately to develop and exercise fully their moral powers and to pursue the determinate conceptions of the good they come to form." PL, *supra* note 16, at 77. The fact that the parties—who, Rawls assumes, are citizens of a well-ordered society—seek principles that will aid the development of their two moral powers makes clear that at the moment of selecting the principles these powers are not at their height.

in the process of selecting the principles of justice, because their basic moral capacities allow them to recognize and to affirm—if not always to act upon—the fair terms of social cooperation.¹⁵⁸

So too, therefore, may we count among the parties selecting the principles of just punishment (most) potential criminal offenders—at least those who, unlike, say, a Charles Manson or a Jeffrey Dahmer, retain some capacity to understand the difference between right and wrong and the obligation of all members of society to act respectfully toward others, even if they sometimes fall short. Indeed, for a partially compliant society, *not* to include members of this group in deliberations over the content of the principles of just punishment would be to rule out consideration of the perspectives of an entire subset of the population before any actions have been committed that might warrant such exclusion. It would thus be to compromise Rawls's own understanding of the nature of this requirement, which he assumes to be "not at all stringent," and to be "possessed by the overwhelming majority of mankind . . ." ¹⁵⁹ To exclude potential criminal offenders on this ground would therefore be to skew the process in a way not at all true to the original intention of the stipulation—and to undermine the potential of the framework to yield insights into the nature of legitimate state power.

Admittedly, as indicated above, it does seem necessary to exclude from participation in this process at least some potential criminal offenders: those individuals—I will call them "sociopaths"—who, like the Charles Mansons and Jeffrey Dahmers of the world, are entirely incapable of dealing with anyone respectfully and want only to violate the fair terms of cooperation with their fellows and do them harm. These are the people Rawls calls "evil," who "aspire[] to unjust rule precisely because it violates what

158. That the parties, and the citizens they represent, will invariably be flawed in these ways only makes sense. There is no point in deriving principles of justice for a society of angels; in such a world, such principles would only be superfluous.

159. TJ, *supra* note 14, at 506.

independent persons would consent to in an original position of equality"¹⁶⁰ Such people are "move[d]" by the "love of injustice." They "delight[] in the impotence and humiliation of those subject" to them, and "relish[] being recognized by them as the willful author of their degradation."¹⁶¹ If we think of members of society as placed along a continuum in terms of their ability to understand and act on fair terms of social cooperation, this subset of evil people would lie at the extreme end of incapacity. And because of their incapacity, we must stipulate that members of this subset are outside the universe of participants in the process of selecting the principles of justice for a shared world.¹⁶²

This is not, however, to exclude individuals in this group from being treated as subjects of justice once the principles of punishment are chosen. Rather, this subset—which we can assume to include Mansons, Dahmers, and the like—must be placed alongside other groups in society which are also, due to incapacity, excluded from the process of identifying the content of the principles. Because the members of these groups—who also include young children

160. *Id.* at 439.

161. *Id.*

162. It might be argued that whether or not one possesses the two moral powers to a sufficient minimum degree to qualify for consideration in the original position is itself morally arbitrary, and thus that denying a Manson or a Dahmer consideration on this basis merely reproduces the unfair exclusions Rawls rejects. But as we have seen, Rawls is not committed to the luck egalitarian position that the purpose of justice is to compensate for the negative effects of unchosen circumstances. He is instead committed to the idea that all those who are capable of recognizing the entitlement of all others in society to consideration and respect as fellow human beings and fellow citizens are themselves entitled to the same measure of consideration and respect in turn. It may be that those few members of society—those "scattered individuals," TJ, *supra* note 14, at 506—that lack the capacity for moral personality are in this position as a result of morally arbitrary contingencies. But they lack it nonetheless and thus cannot be counted on to accept the fair terms of social cooperation with others. For Rawls, this particular incapacity is sufficient to justify their exclusion from consideration behind the veil. Nonetheless, as I go on to note in the text, those individuals who are thereby excluded are still to be treated in accordance with the requirements of justice as determined in the original position. The point is simply that if they are unhappy with the principles, they have no cause for complaint.

and the “seriously injured or mentally disturbed”¹⁶³—lack the requisite capacities for such participation, those parties who do have the necessary capacities are charged with their trusteeship, choosing for them “as we have reason to believe they would choose for themselves if they were at the age of reason and deciding rationally.”¹⁶⁴ The principles of just punishment the parties go on to select will thus bind not only the parties themselves, but also members of the groups for whom the parties have acted as trustees.

Most people who come to violate the criminal law, however, do not fall into the category of sociopath. Instead, most criminal offenders may be expected to recognize the wrongful nature of their crimes, and to know that they have “infringe[d] on the just claims of others” and done them injury.¹⁶⁵ On Rawls’s own terms, therefore, the capacity for this recognition is sufficient to qualify most potential criminal offenders for inclusion and consideration in the deliberative process itself.¹⁶⁶

163. *Id.* at 249.

164. *Id.* at 209. To this, one might argue that sociopaths have a different notion of “rationality,” and thus cannot be reasoned for in this way as if they were, say, children, who when they have grown may be expected to share the perspective of the parties selecting the principles. Rawls, however, stipulates that the parties in the original position are “prevented from knowing any more” about the members of the groups for whom they act as trustees “than they do about themselves, and so in this case they must rely upon the theory of primary goods,” assuming that the good for members of these groups is the same as for themselves. *Id.* This assumption may prove wrong in fact, but it is the principles of justice we are after and not the means to satisfy all desires whatever their content. We may therefore assume that what this process of proxy representation yields is just, both for the parties and for those they represent.

165. *Id.* at 446. See *id.* (suggesting that in feeling guilt “we focus on the infringement of the just claims of others and the injury we have done to them, and on their probable resentment or indignation should they discover our deed”).

166. There is a further possible objection to the inclusion of (potential) criminal offenders in the process of selecting the principles of just punishment: although they may have the capacity for a conception of the good, the particular determinate conceptions that motivate them are patently unjust. For this reason, it may be thought wrong, or at least ill-advised, to admit to the deliberations on an equal footing with their likely victims the perspectives of those who have a preference for unjust and often deeply violative agendas. Admittedly, it does seem right that certain preferences, those hostile to others, should be inadmissible from the start. And in fact, Rawls specifies that conceptions of the good that “are in direct conflict with the principles of justice” may be excluded

2. The Contingencies of Punishment and Crime

There is, however, a second argument for the proposition that the parties could not possibly be uncertain as to whether they might, once the veil is lifted, find themselves to be convicted offenders facing state punishment—that whether one turns out to be a convicted offender facing state punishment is not a function of morally arbitrary contingencies, but is rather the result of having undertaken criminal *actions*, for which one would properly be held responsible. And because the veil obscures only morally arbitrary attributes, and not morally relevant actions—which in any case will manifest only after the veil is lifted and the parties enter society as citizens—the parties deliberating under its effects would have no reason to think that, for morally arbitrary reasons beyond their control, they could wind up convicted offenders facing state punishment.

from consideration and even actively discouraged in a just society. JF, *supra* note 41, at 154. In this category of excluded conceptions, we may readily include criminal conceptions: the preferences of, say, rapists and murderers to rape and murder. But it does not necessarily follow that the perspectives of those who *harbor* such preferences may on this basis be themselves excluded from consideration behind the veil. For even if we were to include such criminal perspectives in the process, there is no reason to fear any corrupting effect, either on the deliberations or on the principles eventually chosen, from the knowledge behind the veil that in society some citizens might hold these illegitimate preferences. For recall that behind the veil, no one is wedded to the promotion of any particular conception of the good, because no one yet knows the particulars of his or her own conception. All the parties know is that, like all citizens, they have *some* such conception; that, once they enter society as citizens, it will be of the utmost importance to them to be able to pursue their particular conception; and that there will inevitably be some unfortunate members of society—at this point, who can say which ones?—whose particular preferences will turn out to conflict with the demands of the principles of justice, and who will therefore be unable to pursue their preferences without risking interference from the state. Even knowing that they themselves might be among this group of unfortunates, the parties would still agree to this result because they would recognize the potential dangers to innocent people—possibly themselves—of allowing certain harmful conceptions to flourish. The framework of the original position thus ensures that the potential corrupting effects of criminal preferences will be nullified, while still affirming the equal moral status of those who might possess them, by allowing the inclusion of their perspective—and consideration of their (legitimate) interests—behind the veil.

This conclusion, however, is too quick. To establish whether the parties could be certain that they would not end up as convicted offenders, we must determine whether, given the information to which the parties have access about their future selves and the society they are to enter, they could be fully confident that they would always be in sufficient control over their actions always to be able to avoid any committing any crimes. For the parties, as they are well aware, are “choosing once and for all the standards which are to govern [their] life prospects” once the veil is lifted.¹⁶⁷ The principles they choose will bind them “in perpetuity[. T]here is no second chance.”¹⁶⁸ If they are to protect their long-term interests, therefore, the parties must be absolutely certain that they consider all possible positions they could turn out to occupy once the veil is lifted. And if they cannot be fully confident that, however things turn out, they will always be able to avoid criminal punishment, they must for their own protection include the perspective of convicted offenders facing punishment among those from which they evaluate the principles of punishment available for selection.

In fact, given the nature and extent of the parties’ knowledge about their future selves and the society they are about to enter, there are three reasons to think that the parties could not be confident that they would always be able to avoid facing state punishment. The first reason arises from the parties’ awareness of the possibility that, once the veil is lifted, they may wind up convicted of—and punished for—crimes they did not commit. The second and third reasons, in a somewhat different vein, reflect the parties’ recognition that, depending on how things turn out, they might well come to be guilty of criminal acts against others, despite understanding full well both the wrongfulness of such acts and the negative consequences that await those citizens who commit them.

167. *Id.* at 176.

168. *Id.*

a. The Threat of Wrongful Conviction

First, under the conditions of a partially compliant society, there is no basis for the parties to be confident that those who will face state punishment as convicted offenders will necessarily be guilty of the crimes charged. To the contrary, in a social context in which the institutions and operations of the criminal justice system are flawed and untrustworthy and known to be so, the parties will know that some (indeterminate) number of convicted offenders incarcerated by the state will in fact be innocent. For in this non-ideal world, even the most honest and upright officials may be expected on occasion to err in ways that create a risk that innocents will be targeted for criminal punishment. And moreover, as the parties are aware, in this society some number of state officials—among them police, prosecutors, defense attorneys, and judges—will at times abuse the considerable powers they have been delegated, thus reinforcing the danger that innocent people will be targeted for investigation, prosecution, and punishment. This subset of officials may be no more than a small fraction of the overall number of state agents charged with the responsibility for administering the criminal justice system. For our purposes, however, all that matters is that the anticipated incidence of wrongful targeting be sufficient to make the parties aware of the real danger that convicted offenders facing state punishment may actually be innocent of the crimes charged.

Given these conditions, which, as stipulated above, are known to obtain in a partially compliant society,¹⁶⁹ it would be a mistake for the parties to assume that only those who commit crimes will wind up as targets of punishment. To the contrary, if the parties are to ensure the greatest protection of their security and integrity, they must instead assume that they could wind up, through no fault of their own, facing state punishment as convicted offenders. The parties selecting the principles of punishment for a

169. See *supra* part III.A.

partially compliant society would thus find themselves on the horns of what we might call the “political dilemma of punishment.”¹⁷⁰ On the one hand, fearing the prospect of victimization at the hands of criminals, they would view the punishment of offenders as a source of protection for their security and integrity.¹⁷¹ But on the other hand, the fear of being wrongly targeted for punishment by state agents would lead them to view the authorization of state punishment as a potential threat to their higher-order interests.

It is of course possible—even probable—that, once the veil is lifted, whether citizens are prosecuted, convicted, and punished by the state for crimes they did not commit will not prove to be random, but will instead prove to correlate with membership in particular social groups. (If we are to take the general facts of human society as our guide, those groups will most likely be those which are

170. Judith Shklar expresses this political dilemma most forcefully, making it the centerpiece of her vision of liberalism, “the liberalism of fear.” Shklar, *supra* note 3, at 21. Shklar takes as her assumption, which she sees as “amply justified by every page of political history,” that “some agents of government will behave lawlessly and brutally in small or big ways most of the time unless they are prevented from doing so.” *Id.* at 28. This assumption is as true of liberal democratic governments as of less freedom-respecting varieties: “public cruelty . . . is made possible by differences in public power, and it is almost always built into the system of coercion upon which all governments have to rely to fulfill their essential functions.” *Id.* at 29. In Shklar’s view, the goal of a liberal state must therefore be to avoid action which is cruel or which generates fear among citizens, fear that is “created by arbitrary, unexpected, unnecessary and unlicensed acts of force and by habitual and pervasive acts of cruelty and torture performed by military, paramilitary and police agents in any regime.” *Id.* at 29; see also *id.* at 27 (“Given the inevitability of that inequality of military, police, and persuasive power which is called government, there is evidently always much to be afraid of. And one may, thus, be less inclined to celebrate the blessings of liberty than to consider the dangers of tyranny and war that threaten it.”). Although she recognizes the necessity of punishment as the only way to prevent “greater cruelties,” *id.* at 30, Shklar insists that “liberalism looks upon [punishment] as an unavoidable evil, to be controlled in its scope and modified by legally enforced rules of fairness.” *Id.* Given the overwhelming coercive power of the state and the tendency to cruelty of all governmental institutions, including liberal ones, Shklar cautions that “any confidence that we might develop in [governmental] agents must rest firmly on deep suspicion.” *Id.*

171. On the question of how the parties may take into account the possible deterrent effects from behind the veil, see *infra* part IV.A.

relatively powerless or politically unpopular.¹⁷²) But although, once the veil is lifted, some citizens might for this reason find themselves insulated from such focused targeting, the prospect of this eventual insulation provides no comfort to the parties behind the veil. For it is precisely the knowledge of the particular characteristics which might ultimately turn out to render one vulnerable to—or protected from—wrongful conviction that the veil of ignorance precludes. And because, as we have seen, the parties lack the information that would allow any estimating of probabilities as to their own chances of being innocents wrongfully targeted for state punishment, the parties must consider what it would mean for *anyone* to be so treated.

b. Humanity's Imperfections and the Inevitability of Moral Error

The parties also have grounds for uncertainty as to whether they could end up, once they enter society as citizens, not as wrongfully convicted innocents, but as guilty offenders facing state punishment.¹⁷³ For although the parties do not know the particulars of their own personal attributes, they *do* know the “general facts about human society,” including the “laws of human psychology.”¹⁷⁴ They therefore know that human beings are not infallible. We make mistakes. We make bad judgments. We act on impulse, and in haste. And often, when we do so, we do wrong to others. Rawls, in constructing his well-ordered society, neatly erases the implications for justice of the facts of human fallibility and

172. See Jim Dwyer et al., *Actual Innocence: Five Days to Execution and Other Dispatches from the Wrongly Convicted* 267 (2000) (indicating that 57% of exonerated defendants are African-American and 11% are Latino).

173. When I speak of guilty offenders in this article, I mean those offenders whose guilt has been adjudicated under procedures consistent with standards of due process and principles of criminal law, akin to those currently governing the American criminal justice system. An exploration of the adequacy or legitimacy of these standards and principles is beyond the scope of this article.

174. TJ, *supra* note 14, at 137.

prone to moral error. He simply assumes that citizens in society strictly comply with the demands of justice and honor their duties and obligations toward their fellows. The parties deliberating for a partially compliant society, however, have no such luxury. Instead, it is a basic premise of the partially compliant society that human beings will sometimes fail to comply with the demands of justice, and will sometimes fail to do right by their fellows. That this is so need not reflect a despairing belief in widespread inhumanity, but simply the recognition of the inevitable limitations and failings of even those human beings who understand the difference between right and wrong and who recognize their obligations toward others.

The parties' awareness of the facts of human fallibility and moral frailty means that, in selecting the principles of punishment to govern themselves for all time, the parties would be unwilling to select principles on the assumption that they will always be in control of their own actions sufficient to avoid committing any wrong that could result in criminal punishment. Knowing the limitations of the real human beings they represent, the parties would be unwilling to approach the selection process as though they would never transgress, and would instead assume that, depending on how things turn out, they could wind up as guilty offenders facing punishment.

*c. The Arbitrariness of Pressures to Offend—and
of the Moral Resources Necessary Always to
Resist*

Finally, the parties, knowing the "general facts about human society" as well as the unjust character of the partially compliant society, would recognize that, depending on the particular social position they turn out to occupy, they could wind up, once the veil is lifted, facing considerable pressures and temptations to offend and without sufficient moral resources always to resist.¹⁷⁵

175. Of the three reasons offered here as to why the parties could not be

Knowing that this is so, the parties would be moved for their own self-protection to consider the perspective of convicted offenders facing punishment when selecting the content of these principles.¹⁷⁶

To show that the parties would be so moved, it is not necessary to demonstrate that the possession of particular (morally arbitrary) qualities would inevitably lead one to violate the security and integrity of others. It would instead be sufficient to demonstrate that the possession of certain morally arbitrary qualities would place those who have them in a position where they could not be confident that they would always be able to resist pressures or temptations to perform such harmful acts. And this latter showing, it seems to me, can indeed be made. To see that this is so, we need only consider a particular range of possible attributes and personal particulars—those combining the effects of social and structural disadvantage with arrested moral development—that would attach to (some) members of a partially compliant society. Nowhere in *A Theory of Justice* does Rawls explicitly consider this particular combination of qualities, but they are nonetheless fully consistent with the

entirely confident of their ability to avoid state punishment once they enter society as citizens, I anticipate that this final reason will be the most contentious. I nevertheless believe it to be sound, for reasons I sketch in the text. Still, it bears noting that, strictly speaking, acceptance of this third and final reason for the parties' uncertainty is not necessary to sustain the broader argument, for the combination of the two reasons already offered is sufficient to lead the parties to consider the interests of convicted offenders in their selection of the principles. This would be so even were the danger of wrongful conviction eliminated, since the inevitability of moral error even among human beings with the capacity for a sense of justice is sufficient to create the possibility in the minds of the parties that they could turn out to be guilty offenders once the veil is lifted.

176. The issue here is not whether those who do in fact go on to offend should ultimately be held responsible for their criminal actions; as we have already seen, on the Rawlsian model, once the parties enter society as citizens, they will indeed be held responsible for any actions that impinge on the goods of others, whatever their reason for so acting. Here, we simply seek to determine the range of perspectives the parties would adopt when determining the principles constraining the consequences for criminal actions which guilty parties will be required to bear.

"certain kinds of particular facts" that the veil of ignorance is intended to obscure.¹⁷⁷

Recall that at this point we seek only those aspects of the person that could properly be understood as morally arbitrary, existing through no fault of the individuals to whom they attach; actions, as we have seen, are not morally arbitrary in this sense and would thus not be obscured by the veil. Still, within this range, it is possible to identify a number of such attributes and personal particulars that could lead their possessors to face strong pressures and temptations to offend, without the moral or other resources necessary always to be able to resist. Consider, for example, the experience of having been born and raised in a family in which physical and psychological abuse was commonplace, in which one learned early to express feelings of frustration, resentment, or anger through violence, and lacked instruction or role models presenting a more respectful and mature way of expressing such feelings.¹⁷⁸ Surely whether or not one's childhood experience takes this form is the function of morally arbitrary contingency. And surely it is not hard to imagine one who was raised in such circumstances finding it tempting and even natural when angry or frustrated to resort to violence against others, and also finding it a much greater challenge to resist this inclination than would those with a more fortunate upbringing.¹⁷⁹

177. TJ, *supra* note 14, at 137.

178. See James Gilligan, *Violence: Our Deadly Epidemic and Its Causes* (1996) (arguing that those who commit acts of extreme physical violence are often motivated to do so by feelings of shame and humiliation that they are otherwise incapable of expressing); Richard Rhodes, *Why They Kill: The Discoveries of a Maverick Criminologist* 112-24 (1999) (describing sociologist Lonnie Athens's findings that extremely violent people have uniformly had the early experiences of being forced through violence or threat of violence to submit to the will of another, and of watching loved ones be similarly treated).

179. To say this is not to excuse the actions of one who, having grown up under such circumstances, uses violence against others, nor to imply that such a person ought to escape punishment. It is simply to explain how it might be that one's attributes and characteristics, developed as a result of morally arbitrary contingencies, might lead one to find oneself in circumstances in which it is extremely difficult to resist the impulse to do harm to others.

Or consider the experience of being born into a very poor family, lacking access to an adequate education, raised by parents forced to work long hours just to get by and thus unable to provide their children with either supervision or moral guidance. Under these (morally arbitrary) circumstances, one might find few opportunities for developing skills or otherwise breaking out of the cycle of poverty and violence found in one's neighborhood. Indeed, one might find—through no fault of one's own—that illegal activities provide the only means for economic security or structured occupation.¹⁸⁰ Or imagine finding oneself, again as a result of morally arbitrary contingencies, part of a racial or ethnic or religious group that is systematically discriminated against by the broader society. Denied access to decent schools or other opportunities and constantly humiliated by or made to feel inferior to members of the majority, one would face considerable obstacles to developing as a mature and confident person able to build a satisfying life and deal with others in a healthy and respectful manner.¹⁸¹ And imagine, in all of these cases, having had as role models only those who have themselves taken the path of criminal activity and even violence, perhaps having themselves spent time in prison.

Under conditions of partial compliance, these imaginings represent the circumstances the parties might well face in their own lives once the veil is lifted. The question then becomes: how, if at all, would recognition of these possibilities on the part of the parties affect their thinking as to whether they could turn out to be guilty offenders facing punishment? Given the nature and range of these morally arbitrary positions, it is simply not

180. See, e.g., William J. Wilson, *When Work Disappears: The World of the New Urban Poor* (1996); Recent Legislation: Welfare Reform—Punishment of Drug Offenders—Congress Denies Cash Assistance and Food Stamps to Drug Felons—Personal Responsibility and Work Opportunity Reconciliation Act of 1996, Pub. L. No. 104-193, § 115, 110 Stat. 2105 (to be codified at 42 U.S.C. § 862A), 110 Harv. L. Rev. 983 (1997).

181. Bigger Thomas, the anti-hero of Richard Wright's novel *Native Son*, is a particularly poignant and disturbing portrait of the possible effects of such circumstances. See Richard Wright, *Native Son* (1940).

realistic to assume that the parties would be confident that no matter how things turned out for them, they would always be able to abstain from committing violent or otherwise harmful actions against others. This is particularly true given the volatility, the constrained options, and the limited moral and other resources that would define the lives and possibilities of those finding themselves in the circumstances here described.¹⁸²

To this suggestion, it might well be objected that the parties, with their two moral powers, would have greater confidence than I suggest in their ability to resist any pressures or temptations to do harm to others. But recall that the parties are to select principles of justice to bind them for the long term. To be certain that they would never find themselves targeted for state punishment as guilty offenders, the parties must be confident, not that they might *at times* be able to resist any pressures or temptations they might face to offend, but that they will *always* be able to do so. And under the circumstances of the partially compliant society, mindful of the strains of commitment, the parties simply could not be confident in this regard.

Why not? As our discussion has already suggested, in a partially compliant society, whether one is able to develop the moral and cognitive capacities of a mature, confident person, able to exercise judgment and self-control and to deal with others in a respectful manner, is itself a function of morally arbitrary contingencies. Deliberating behind the veil, the parties thus could not be confident that once the veil is lifted they would necessarily turn out fortunate enough to develop the moral resources necessary to ensure

182. Again, my claim is not that people with these experiences would necessarily turn to crime. It is simply that, in a partially compliant society, there will be some people whose circumstances create for them greater pressures and temptations to do harm to the security and integrity of others than do the circumstances of more fortunate members of society. Admittedly, I have here offered extreme examples, but given the conditions of partial compliance, they are by no means fanciful. To see that this is so, one need only read the life histories of many of the men and women in prison or on death row across the United States.

that they will never act in ways that violate the security and integrity of others. For this reason, even absent any danger of wrongful conviction, the parties would still perceive a possibility that they could wind up guilty offenders facing punishment, and would thus still—to protect their own long-term interests—consider the implications of proposed principles for those targeted for punishment by the state.¹⁸³

D. The Rawlsian Model and the Managing of Emotions

In this part, I have argued that, assuming the conditions of a partially compliant society, the parties would face uncertainty as to whether, once the veil is lifted, they will be able to avoid occupying the positions of crime victim or target of state punishment. In order to ensure the greatest protection possible for their security and integrity, they would therefore be sure to consider how

183. This claim may seem to fly in the face of Rawls's insistence that all members of society have, as part of their capacity for a conception of the good, the capacity to revise their ends when acting on those ends would burden the legitimate expectations of others. Rawls, however, assumed this capacity to exist as part of a broader social division of responsibility. On the one hand, "citizens as individuals accept responsibility for revising and adjusting their ends and their aspirations in view of the all-purpose means they can expect, given their present and foreseeable situation." On the other hand, society, the collective body of citizens, "accepts the responsibility for maintaining the equal basic liberties and fair equality of opportunity, and for providing a fair share of the other primary goods for everyone within this framework." SU, *supra* note 92, at 371. Where society is defined by the conditions of partial compliance, however, it has not lived up to its part of the bargain. In a partially compliant society, we know that not all citizens abide by the duties and obligations they owe to others. And we also know that society's background conditions are unjust and known to be so. Certainly, under these conditions, we can still expect citizens to have a general capacity to form, revise, and pursue their own conception of the good. No actually existing society has ever achieved the conditions of Rawls's well-ordered society, yet people existing in all manner of imperfect societies continue to demonstrate the human capacity to develop and pursue their own aims and plans despite broader conditions of deprivation and injustice. What we cannot reasonably expect of citizens in a partially compliant society is the capacity always to be able to revise their ends in order to resist acting in ways that violate the security and integrity of others. Instead, the best we can hope for is to keep failures in this regard to a minimum. To do so is the motivating aim of the parties selecting the principles of punishment. For discussion, see *infra* part IV.

alternative principles of punishment would impact the security and integrity of both these social positions—and to select those principles that would maximize the prospects of the occupants of whichever position turns out to be the worst-off in this regard. The parties will thus accord due consideration to the interests of all citizens in the selection of the principles of punishment, whatever they may come to think about actions and choices of particular individuals once the veil is lifted.

To some readers, the foregoing may seem an unnecessarily complex and cumbersome way of reaching this point. If we want this range of perspectives represented in the process of selecting principles of punishment that everyone subject to them would agree are just and fair, it might be thought, why not just gather together a group of people who are trusted and respected and exhort them to select such principles, according due consideration in their deliberations to the perspectives of all members of society, including that of criminal offenders?¹⁸⁴ Not only would such a straightforward approach be much simpler, but it would avoid the arguable “perversity” that attends Rawls’s approach, which models “the idea that other people matter [as ends] in . . . themselves, not simply as a component of our own good,” by requiring the parties to consider and care about only what is in their own self-interest.¹⁸⁵

184. And if they balk at according such consideration to a Charles Manson or a Jeffrey Dahmer, we can simply create an exception matching the above exclusion of the interests of sociopaths from the process.

185. Kymlicka, *supra* note 41, at 69. As Kymlicka puts the point, [T]here is a curious sort of perversity in using the contractarian . . . device to express the idea of moral equality. The concept of a veil of ignorance attempts to render vivid the idea that other people matter in and of themselves, not simply as a component of our own good. But it does so by imposing a perspective from which the good of others is simply a component of our own (actual or possible) good. The idea that people are ends in themselves gets obscured when we invoke “the idea of a choice which advances the interests of a single rational individual for whom the various individual lives in a society are just so many different possibilities.” . . . Rawls tries to downplay the extent to which people in the original position view the various individual lives in society as just so many possible outcomes of a self-interested choice, but the contract device

Yet however well such a simple and direct approach might work under assumptions of full compliance and fair distributive justice—and given the human tendency toward self-interest and the way this tendency can distort our sense of what is just and fair even under relatively ideal conditions, the challenge would be great even assuming the background conditions of a well-ordered society—it could never adequately serve when our aim is the identification of principles of punishment. For the arena of crime and punishment is just too charged, too fraught with emotions—hatred, resentment, contempt, scorn, outrage, fear—to allow personally involved participants to retain the level of detachment required to achieve the necessary measure of impartiality. Such emotions are understandable, even appropriate, when directed against criminal offenders, violent offenders in particular. But when not carefully managed, they can easily create an atmosphere in which it would be impossible to accord the interests of all members of society, criminal offenders included, the kind of consideration that would be required to yield principles of punishment that all members of society, including those subject to them, could agree are just and fair.

As we will see below, according such consideration does not require the conclusion that criminals and their victims are ultimately entitled to equal treatment. To the contrary, even if the parties, in order to protect their own interests, were to consider the effects of proposed principles on convicted offenders, they would nevertheless readily adopt principles that authorize the imposition of punishment—at times severe punishment—on members of this group.¹⁸⁶ But according such consideration to all members of society *would* require some measure of dispassion sufficient to recognize

encourages that view, and so obscures the true meaning of equal concern.

Id.

186. As we will see, they would do so because, as potential crime victims themselves, they would recognize the wrong done by criminals when they violate the security and integrity of their victims, and thus the justice of imposing punitive hardships on convicted offenders who have done wrong to others when doing so would protect the security and integrity of law-abiding citizens.

and consider the extent to which the interests of criminal offenders should come into play in the shaping of the principles. And it is precisely because such dispassion is so difficult to sustain in the real world that the Rawlsian construction is so useful to us. For if anything is to help us gain perspective on whether particular criminal punishments are actually just and fair to all concerned, it is a construction of the issue that forces us to consider, not what "those others" deserve, but what measure of punishment we ourselves would accept as justified if we were the people branded as criminals.

By now, it should be clear why I have constructed the conditions of the partially compliant society as I have: these conditions closely resemble those of our own social context. Although it might be thought that a focus on this vision of society will yield principles that are merely contingent on continued problems with criminal justice enforcement and unfavorable background conditions, such a notion, it seems to me, reflects an unduly optimistic picture of the potential of actual liberal democratic societies either to rein in state agents' tendency to overreach or to alleviate distributional injustice. Of these two possibilities, rendering the criminal justice system sufficiently trustworthy seems the less utopian goal of the two. In part IV, I thus consider where relevant how, if at all, the principles might change were the danger of wrongful conviction eliminated. Still, I think it noncontroversial (if unfortunate) to assume that all three conditions of the partially compliant society represent permanent features of any foreseeable liberal democratic society.¹⁸⁷ For this reason, I consider this model to be that on which we should focus our efforts to identify the principles of legitimate punishment for such societies—including our own.

187. See Shklar, *supra* note 3, at 27 ("Given the inevitability of that inequality of military, police, and persuasive power which is called government, there is evidently always much to be afraid of [even in a liberal society]."). See also *id.* at 28 ("The assumption, amply justified by every page of political history, is that some agents of government will behave lawlessly and brutally in small or big ways most of the time unless they are prevented from doing so.").

There is, however, a second and more immediate reason to consider what the content of the principles of punishment would be were the dangers of wrongful conviction eliminated. To the extent that the parties' willingness to consider the perspective of targeted offenders is thought to derive solely from their uncertainty as to whether they might turn out, once the veil is lifted, to be wrongfully convicted offenders, one might be tempted to conclude that the protections the principles thereby accord to targeted offenders no longer hold in cases in which the guilt of the offender is incontrovertible. As I show, however, this conclusion is not borne out by the analysis. It is to underscore this point and to make clear the extent to which the parties would even on this more limited perspective extend some consideration to the interests of guilty offenders that, where relevant, I highlight in the discussion below the nature of the differences and similarities between the principles the parties would select under the full conditions of partial compliance and those they would select were the danger of wrongful conviction eliminated.

IV. DERIVING THE PRINCIPLES

Under the full conditions of partial compliance, the parties will thus recognize three undesirable possibilities in terms of the social positions they could turn out to occupy once the veil is lifted: crime victim, wrongfully convicted innocent, and guilty offender. What is it about each of these positions that renders them undesirable? The parties know that, once the veil is lifted, they will find themselves with their own particular conception of the good. Behind the veil, the parties cannot know the details of the particular conception of the good they will ultimately hold, but they do know that it will be important to them to enjoy the freedom to pursue it without interference, whatever it turns out to be. And, as we have seen, to ensure this freedom, they need most urgently the set of goods to which

I have referred collectively as “security and integrity”.¹⁸⁸ security from assault on and interference with their physical and psychological integrity and well-being, and protection against any who would compromise these goods or put them at risk. Yet recognizing that the threat to their security and integrity could come both at the hands of fellow citizens and at the hands of the state, the parties would face what I have called the political dilemma of punishment. Fearing crime at the hands of their fellows, the parties would look to the state to impose punishment as a way of providing them protection. At the same time, fearing the violation of their security and integrity that punishment represents, they would be wary of investing the state with the authority to exercise this kind of coercive power. The challenge the parties thus face is to identify principles of punishment that would answer both concerns, providing them the protection they will need in a partially compliant world both from fellow citizens and from the state.

A. The Lexical Difference Principle and the Deterrent Effect

One may wonder why the parties, uncertain of their social position in the ways we have noted and concerned exclusively to maximize the prospects of the worst-off in society, would authorize any punishment at all. For reasons we will shortly see, the parties would ultimately reject a “no-punishment” principle. Still, it is worth examining the argument that might support such a principle, for doing so exposes the way the parties, despite being unable to estimate probabilities, are still able to select alternatives based on their effects on crime rates overall.

The argument for a no-punishment principle might run as follows. Assume the commission of a violent crime,

188. The parties would also recognize the importance of a basic minimum of material resources to the possibility of pursuing their particular conception of the good. As I have argued, however, security and integrity are prior even to the need for material resources. See *supra* pp. 353-55.

say, a stabbing. From the perspective of the parties, there would be two alternative regimes of punishment with which society might respond to a battery of this sort, neither of which would satisfy the demands of maximin. On the first, the convicted offender is sentenced to a punishment that puts him or her in a worse position than that occupied by the victim. But because the parties judge principles solely on the basis of their impact on the worst-off in society, they would reject this first option.¹⁸⁹ They would do so because this option creates the possibility of a fate even worse than that of stab victim, that of punished offender, and the parties—who fear winding up either crime victim or targeted offender—would presumably prefer a regime that did not create a worse worst-off position to one that did.

The second alternative—that the convicted offender is sentenced to a punishment which puts him in the same (undesirable) position as the victim,¹⁹⁰ or at least in a less desirable position than the offender would otherwise have occupied absent punishment—would, on this reasoning, be equally unappealing to the parties. For on the terms we have stipulated, there will in any case already be one person (the victim) in the worst-off position, and the parties, reasoning according to maximin, would not want to endorse an alternative that would place still another person (the convicted offender) in a less desirable position than they would otherwise occupy. Presumably, although the parties cannot calculate probabilities, each knows that he or she can occupy only one position in this scenario.¹⁹¹

189. Recall that the parties regard both the position of crime victim and the position of convicted offender as positions they could end up occupying. They therefore seek to maximize either position should it come to represent society's worst-off.

190. There are, of course, significant obstacles to any attempt to design punishments that are exactly proportionate to the harm caused by the crime. See David Dolinko, *Three Mistakes of Retributivism*, 39 *UCLA L. Rev.* 1623, 1636-1642 (1992). But because I suggest this possibility merely for the sake of argument and not as an assertion regarding its empirical possibility, these practical difficulties pose no problems for the present discussion.

191. For purposes of this argument, I leave aside the complicated scenario in which each party at the scene of the crime is both assailant and victim.

The parties would thus prefer an outcome with fewer worst-off positions, and would therefore on this reasoning also reject this latter approach to state punishment.

This analysis seems consistent with maximin as we have explored it. The obvious difficulty, of course, is that viewed in this way, maximin reasoning seems to preclude the parties from accessing the deterrence potential of punishment—a potential that the parties would very much want to tap, given their primary interest in protecting their security and integrity. This reasoning, moreover, requires the parties to consider appropriate punishments for crimes viewed in isolation, as if punishments for one crime would have no effect on the crime rate overall. Yet the parties, aware of the “general facts of human society” and the “laws of human psychology,”¹⁹² would surely recognize that without some disincentive to criminal behavior, more crimes against citizens will be committed than would be committed were such a disincentive actually in place.¹⁹³

As we have seen, however, the parties have no basis for estimating probabilities, a constraint which is necessary to ensure that the parties attend to “all of the lives under consideration,” whether or not they turn out to be their own or someone else’s.¹⁹⁴ Yet this capacity would seem a necessary component of recognizing and weighing the deterrence value of punishment. Is this then to say that, despite their likely awareness of the relationship between disincentives to crime and the rate of its commission, and despite moreover their reasonable preference for a society with less crime rather than more, the parties would be unable to reason about punishment in ways that take account of its deterrent potential? For, if this were so, it

192. TJ, *supra* note 14, at 137.

193. To make this assumption, we need not claim that such disincentives will ever be 100% effective, nor that all people in society will respond to them equally. For our purposes, the claim need only be the very modest one that, absent the threat of some punishment—some treatment “involv[ing] pain or other consequences normally considered unpleasant,” H.L.A. Hart, *Punishment and Responsibility* 4 (1968)—crime rates would be higher than they would be were such a threat known to be present.

194. Hurley, *supra* note 105, at 382.

would be difficult indeed to imagine that the Rawlsian framework could yield principles of punishment citizens would willingly affirm.

Fortunately, this worry turns out to be unfounded. To see why this is so, consider two states of affairs. On the first, n people are victims of armed robbery, leading n people to suffer the violation of their security and integrity this crime represents. On the second, $n - y$ (where $n > y$, and y is a positive number representing the number of people who were victims of armed robbery on the first scenario who were *not* victimized on the second due to the successful deterrence of their assailants) people are victims of armed robbery and thus suffer a like compromising of their security and integrity. I assume here that, under conditions of partial compliance, we will never achieve perfect deterrence—meaning that, even if we are able to deter some criminal activity, we will never successfully reach a point at which no incidents of any given crime will take place. But even recognizing that, under prevailing conditions, there will always be some victims of armed robbery (in other words, that n will always be greater than y), the parties may nonetheless be expected to prefer the second state of affairs to the first, all other things being equal. And, for reasons Rawls captures in what he calls the “lexical difference principle”¹⁹⁵ (“leximin”), they may do so consistent with the maximin reasoning that, as we have seen, the parties would necessarily adopt behind the veil.

The idea of leximin is a simple one: when, as here, no available course of action will improve the circumstances of the worst-off person (because $n > y$ and may be expected always to be so), the course should instead be adopted, assuming one is available, that will improve the prospects of the next worst-off person whose circumstances stand to be positively affected.¹⁹⁶ Although the parties will be in no

195. TJ, *supra* note 14, at 83.

196. As Rawls explains it:

[I]n a basic structure with n relevant representatives, first maximize the welfare of the worst off representative man; second, for equal welfare of the worst-off representative, maximize the welfare of the second worst-off

position to know in advance how, if at all, this approach will benefit them personally, they need not have this information to opt for it. For where the circumstances of the worst-off person will remain unchanged, the parties may still best protect their interests by improving the lot of the next worst-off person who stands to be affected.¹⁹⁷ In the context of punishment, this possibility would allow them at least to consider the option of improving the lot of some citizens by reducing the incidence of crime with the threat of punishment—a reduction we will call “the deterrent effect”—even knowing as they do that at least some citizens (perhaps themselves) will still wind up as victims. This deterrent effect may be the product of specific deterrence, general deterrence, or incapacitation. All that must be shown to satisfy the requirement of the deterrent effect is that the threat of punishment reduces the incidence of crime.

But is the introduction of *leximin* enough to render the authorization of state punishment an appealing prospect from the perspective of the parties? Certainly, under circumstances in which the parties would consider the question solely as future victims, the answer is yes: the parties could be expected to reject a no-punishment principle on the basis of *leximin*, so long as the punishment scheme adopted would be expected to yield $n - y$ crime victims (where y represents what we can now recognize as

representative man, and so on until the last case which is, for equal welfare of all the preceding $n-1$ representatives, maximize the welfare of the best-off representative man. We may think of this as the lexical difference principle.

Id. at 83.

197. Notice that, for Rawls, the identification and selection of the difference principle and its complicating analogue, the lexical difference principle, marks the conclusion of the process. Because on his scheme the parties are selecting principles of distributive justice for a well-ordered society, they need not fear, as we do, alternatives that worsen the situation of any party. This potential worsening of the circumstances of particular citizens is a complicating feature of our endeavor and means that, unlike on Rawls's account, more must be said to determine the content of principles of punishment to which the parties would agree. The lexical difference principle is nonetheless an important step along the way to such a determination.

the deterrent effect) rather than n such unfortunates. And, as we will see, the same holds true even were the parties to include the perspective of guilty offenders in their calculus.¹⁹⁸ Once the possibility of wrongful conviction is introduced, however, the parties may be expected to reach the opposite conclusion, at least in one case: where the system is so untrustworthy that the danger of wrongful conviction is thought to be greater than the likelihood of one's benefitting from the deterrent to crime.¹⁹⁹ For assuming such an untrustworthy system, the parties would come to view state punishment itself as the greater threat, worsening the position of targeted citizens without thereby producing sufficient countervailing benefits to potential crime victims otherwise worse off. It is thus at this point, at which the danger of wrongful conviction is thought to be greater than the benefit from deterred crime, that the parties would reject the institution of punishment altogether;²⁰⁰ punishment under these circumstances would violate the demands of *leximin*, which requires that, all other things being equal, no one be made worse off if such a result can be avoided by an alternative approach.

Thus, if the parties are to opt for any punishment principle at all in a system prone to wrongful conviction, we must assume that, even under the non-ideal conditions of partial compliance, the system is not so untrustworthy that

198. Considering this broader perspective requires a more complicated formula (where the imposition of punishment may be expected to yield $n - y$ crime victims, and where $y > 1$). I therefore defer discussion of this more complicated question to part IV.B below.

199. In the context of such an untrustworthy system, punishment would in any case achieve little deterrent effect. For the disincentive effect of punishment requires potential wrongdoers actually to believe that negative consequences will attend their actions. And if citizens sufficiently doubt the credibility of the determinations of guilt reached by the criminal justice system, they will no longer perceive the connection between wrongful conduct and punishment that is necessary for the threat of punishment to reduce the number of overall victims from n to $n - y$.

200. See TJ, *supra* note 14, at 241 ("The establishment of a coercive agency is rational only if these disadvantages are less than the loss of liberty from instability.").

the threat of state punishment achieves no appreciable deterrent effect.

B. Punishment in a Partially Compliant Society

Armed with leximin and an understanding of the parties' perspectives on the possibilities they face behind the veil, we are now in a position to determine the content of the principles the parties would select, assuming the conditions of partial compliance. Although we have identified three possible positions the parties could turn out to occupy once the veil is lifted—crime victim, wrongfully convicted innocent, and guilty offender—the parties facing these conditions need in fact only consider the issues before them from two perspectives, those of crime victim and of convicted offender facing punishment. For where the system is known to target innocents along with the guilty, it is impossible to know who among the punished is guilty and who is innocent. Considering the interests of the wrongfully convicted will therefore necessarily extend the same consideration to the guilty. Once we have identified the principles the parties would adopt on the full conditions of partial compliance, however, we will also be in a position to see the difference it would make to the conclusions reached were the system reformed to eliminate the danger of wrongful conviction,²⁰¹ leaving the parties confident that all convicted offenders are in fact guilty.

Assuming the full conditions of partial compliance, then, what principles of punishment would the parties adopt?

201. Any system of punishment, no matter how effective the safeguards, will inevitably punish some innocents along with the guilty. For this reason, it might be thought that there would never be any need for the principles the parties would choose were the danger of wrongful conviction eliminated. Nonetheless, for the reasons noted above, see *supra* pp. 377-78, I believe it instructive to consider this question, even if the inevitable residual risk of wrongful conviction present in even the most reformed system would arguably be enough to drive the maximin reasoning in the same direction it takes when we assume the full conditions of partial compliance.

1. Punishment of Non-Serious Offenses

Call any action on the part of a citizen which credibly and seriously²⁰² violates or threatens the security and integrity of another person a "serious offense."²⁰³ Given this definition, it follows that a "non-serious offense" is an action that does *not* credibly and seriously violate or threaten the security and integrity of another person. In many cases, the designation of a given offense as serious will not be a matter of much contention. Murder, rape, armed robbery, kidnapping: categorizing these criminalized behaviors as serious offenses as here defined would likely garner no objection. The designation of a given offense as non-serious, however, is another matter. Does shoplifting credibly and seriously violate or threaten the security and integrity of another person? Does prostitution? Drug possession? Painting graffiti on private property? Stealing a car radio through an open car window? Whether or not these or other arguably eligible offenses would in fact qualify as non-serious would be open to debate.²⁰⁴ It is for this reason that I make no attempt here to classify the range of existing offenses in terms of their status as serious or non-serious. Instead, I merely posit the existence in a partially compliant society of both serious and non-serious offenses, leaving the precise itemization of which offenses fall within this category to policymakers at the legislative stage.²⁰⁵

202. I use the words "credibly" and "seriously" to indicate that the threat or harm must be more than *de minimis* and that it must be such as would be so construed by a reasonable person in the victim's position.

203. On this definition, it bears noting, the category of serious offenses would encompass both violent crimes against the person and certain crimes against personal property.

204. There is no danger that this category will turn out to be a null set, as theoretically, barring any constitutional problems, *any* action that might be undertaken could be labeled by a legislature as a criminal offense. The present analysis simply provides a way to understand the scope of punishment the state may legitimately impose for the commission of those offenses that are non-serious on the definition I have here stipulated.

205. On the legislative stage and what is to occur there, see *infra* part V.B-C.

Assume, therefore, the commission in society of both serious and non-serious offenses. Assume further (at least for the moment) that the only available mode of criminal punishment is incarceration under humane conditions for a specified term,²⁰⁶ of a length sufficient to represent a serious compromising of the security and integrity of the incarcerated.²⁰⁷ These twin assumptions allow the identification of the first principle of punishment the parties would adopt. That is, under the conditions of partial compliance and assuming incarceration under humane conditions to be the only available punishment, the parties would in the ordinary case²⁰⁸ refuse to authorize the imposition of state punishment for non-serious offenses.²⁰⁹

The reasoning by which they would reach this conclusion is as follows. As we have seen, all parties seek more than anything the protection of their security and integrity. To the extent that incarceration represents a "drastic" interference with the security and integrity of the incarcerated,²¹⁰ the parties would thus consider authorizing the imposition of such punishment on any member of society only if, by so doing, either the occupant of the worst-off position would see an improvement in terms of his or her security and integrity, or, assuming the worst-off

206. For discussion of the question whether the parties, considering the question from behind the veil, would authorize the state to incarcerate convicted offenders under less congenial circumstances, see *infra* part IV.E. See also *supra* p. 325 and note 38, explaining the exclusive focus of the present inquiry on the punishment of incarceration.

207. The precise determination of this length of sentence would be a question to be resolved at the legislative stage. See *infra* part V.B-C.

208. The parties would arguably recognize an exception to this principle in cases where the imposition of punishment would result in an appreciable deterrent effect against the commission of serious offenses. On the "appreciable deterrent effect," see *infra* pp. 392-93. On the elaboration of this exception, see *infra* part IV.B.4.

209. This conclusion would not necessarily preclude some other remedy or social response to non-serious offenses. Because I am restricting my analysis to the punishment of incarceration, however, I will not explore such alternative responses here.

210. See TJ, *supra* note 14, at 242 (noting that imprisonment is "a drastic curtailment of liberty"); JF, *supra* note 41, at 47.

position remains unchanged in this regard, the occupant of the next worst-off position whose security and integrity stands to be affected would see such an improvement. Yet were incarceration authorized for individuals convicted of non-serious offenses, neither of these conditions would ordinarily obtain. As we have just seen, non-serious offenses by definition do not violate or threaten the security and integrity of the victims. Thus, even assuming that the incarceration of non-serious offenses would deter the commission of other non-serious crimes, the beneficiaries of this effect would experience no change in the extent of their security and integrity. They would only thereby enjoy benefits of a less urgent kind—an exchange that, given the priority of the goods of security and integrity, the parties would not agree to make. At the same time, however, the security and integrity of the punished offender *would* be diminished as a result of his or her incarceration²¹¹—and, depending on the severity of the sentence, potentially diminished considerably. The parties, not knowing what position they will occupy once the veil is lifted, but knowing that they could end up as crime victim or punished offender, would thus not endorse incarceration for non-serious offenses. For the only effect of such authorization would be to worsen the circumstances of the worst-off on the metric that matters most—that of the protection of citizens' security and integrity—without any like benefits to anyone.

This conclusion would hold, furthermore, even in cases where the non-serious offense is one that arouses a strong emotional reaction in fellow citizens, prompts widespread feelings of great indignation or contempt, or otherwise grounds fervent and widely-held desires that the guilty

211. It does not matter to the analysis whether we understand a convicted offender to be (1) in the worst-off position in terms of this particular transaction (as compared, for example, with the victims or potential victims of the offense of conviction), thus satisfying the demands of simple maximin; or simply (2) the worst-off person who stands to be affected, the welfare of all other worse-off people remaining unchanged as a result of the punishment, thus satisfying the demands of leximin. Either way, the conclusion remains the same.

offender be severely punished. For in such a case, the incarcerated offender would suffer the violation of her security and integrity, while the onlooker, absent the imposition of punishment, would only face an ungratified desire that the offender be punished. By comparison, the offender would be in the worst-off position. And given what is at stake for the offender, the parties (who would view both the punished offender and the onlooker as positions they could end up occupying once the veil is lifted), would not agree to a scheme that imposed punishment in this case. Considered from behind the veil, the "cruel delight"²¹² that some might derive from watching others suffer harsh punishments, however strongly felt, would thus not be accepted as adequate justification for the punishment of non-serious offenses—or indeed, on this same reasoning, for any offense.²¹³

Certainly, the parties' interest in the satisfaction of their feelings of *Schadenfreude* must be distinguished from the feelings of affirmation crime victims enjoy when the state punishes the perpetrator of the offense against them. To the extent that this affirmation would contribute to the restoration of the victim's sense of security and integrity, the parties would view the benefit to the victim of punishing the guilty offender as a justifiable interest in punishment. In the case of non-serious offenses, however, where there has been no appreciable violation of the victim's security and integrity, there would be no potential for the punishment of the offender to affirm such a violation. The desire for such affirmation would thus not justify a punitive response by the state to non-serious offenses.²¹⁴

212. David Garland, *Punishment and Modern Society: A Study in Social Theory* 63 (1990); see also John Portmann, *When Bad Things Happen to Other People* 138 (2000) (explaining that, for Nietzsche, "[t]he legal institution of punishment amounts to a muted festival of cruelty").

213. See TJ, *supra* note 14, at 230 ("[W]henever questions of justice are raised, we are not to go by the strength of feeling but must aim instead for the greater justice of the legal order.").

214. But see *supra* note 209 (noting that this rejection of the punishment of incarceration for non-serious offenses would not rule out other social responses to

The parties would, moreover, adopt this principle forbidding punishment for non-serious offenses not only under full conditions of partial compliance, but also under conditions in which the danger of wrongful conviction had been eliminated and the parties could therefore assume the guilt of all punished offenders. True, as we will see, where the parties can be confident in the guilt of convicted offenders, they would be willing to authorize punishments that compromise the target's security and integrity where doing so would improve the security and integrity of even one law-abiding citizen. In the case of non-serious offenses, however, punishment of the offender would compromise the security and integrity of the punished while yielding no countervailing benefit in terms of the security and integrity of the law-abiding.²¹⁵ The parties would thus adopt a principle forbidding incarceration for non-serious offenses, even where the offenders are known to be guilty as charged.

2. Punishment of Serious Offenses

The punishment of serious offenses presents a more complicated case. Assuming the full conditions of partial compliance, and assuming also an appreciable deterrent effect²¹⁶—that is, a deterrent effect of y where $y > 1$ —the parties would certainly authorize some measure of punishment for serious offenses, at least to the extent that the punishment compromises the offender's security and integrity no more than the crime itself compromised the

their commission).

215. This is the definition of non-serious offenses stipulated above. Where the victim has suffered such a harm, the offense must be classified as serious.

216. For fuller explication of the appreciable deterrent effect, see *infra* pp. 392-93. The idea here is that the punishment of a given offense should deter at least two future serious offenses. As noted above, this standard could be satisfied by a showing of deterrence in any of its forms, including general deterrence, specific deterrence, and incapacitation. Admittedly, applying a standard like this one raises difficult and important empirical questions, including: Who makes this judgment? On the basis of what evidence? Using what evaluative standards? I address these questions below. See *infra* part IV.B.6.

security and integrity of the victim.²¹⁷ The more difficult question is whether there are any circumstances under which the parties would allow punishment *more* severe in its compromise of the offender's security and integrity than that caused by the offense. For reasons we will shortly see, I believe the parties would authorize disproportionately severe punishment under certain circumstances. But before we can see why this is so, we must first understand the reasons why the parties would readily authorize the imposition of punishment on serious offenders at least where the punishment is of a severity equal to or lesser than that caused by the crime.

Both the commission of serious criminal offenses and the punishment of incarceration negatively impact the position of targets measured in terms of their security and integrity. The parties would recognize that, absent such punishment, there would be no deterrent effect and thus more victims of serious offenses overall. And although punishment of an equal severity to the crime would mean that the position vis-à-vis the security and integrity of the punished offender would sink to the level of his victim, the parties would also recognize that such punishment would, assuming an appreciable deterrent effect, improve the position of a number of other citizens, who would otherwise suffer from undeterred crimes.²¹⁸ As a result, no party would, through the imposition of punishment, be made to occupy a position worse than that occupied by the worst-off person (in this case, the original victim). At the same time, a number of people who would otherwise have been victims

217. Such a scheme would at some point require a determination as to whether the harm to the security and integrity of the victim should be judged based on an objective standard (such as a "reasonable person" could be expected to suffer) or a subjective standard (taking the particular circumstances of the victim into account). My own sense is that for determinations of this sort a more objective standard may be appropriate, although a demonstration as to why this is so is beyond the scope of this paper.

218. The precise number of other potential victims whose lot would be improved would be unknown to the parties, but as long as we stipulate that it must be more than the number of offenders punished, the parties would endorse this reasoning.

of an equally serious crime and thus come in this way to occupy the worst-off position would see their position improve. Thus even though at least two parties now occupy the worst-off position (the original victim and the offender suffering a punishment of equal severity to the crime itself, along with any other victims whose assailants were not deterred), the failure to punish the crime where an appreciable deterrent effect *could* have been achieved through punishment would necessarily mean more victims, and thus more than this number occupying the worst-off position. Because this latter result would violate leximin by failing to maximize the position of the next worst-off position that could have been improved through punishment (victims of like crimes whose assailants might have otherwise been deterred), the parties would instead authorize the other alternative on the table: punishment equal to or less than the severity of the harm caused by the offense.²¹⁹

This result follows from the definition of the appreciable deterrent effect, which is the standard that punishment must be shown to meet before the parties would authorize it under the full conditions of partial compliance. This standard is met where the punishment of offenders would reduce the victims of serious offenses from n to $n - y$, where y is a positive number signifying the deterrent effect, $n > y$, and $y > 1$.²²⁰ This last stipulation—that $y > 1$ —means that the punishment of a convicted offender must ensure that at least two future serious offenses must be deterred, thus saving two potential

219. It bears noting that the quantum of punishment that would be authorized on the above reasoning would in many cases be considerable: it would, for example, arguably allow sentences of life in prison without parole in any case of murder or other crimes which permanently and severely debilitate the victim—at least where such a severe punishment may be shown to be reasonably certain to have an appreciable deterrent effect. Where, as here, incarceration is the only available punishment, life in prison without parole under humane conditions of confinement is the most severe sentence available. For a discussion of the willingness of the parties deliberating under conditions of partial compliance to approve punishments involving more severe conditions, see *infra* part V.E.

220. See *supra* p. 390.

victims from harm. For were punishment of the convicted offender authorized where it would only deter one other serious offense and thus save one potential victim from harm, the punishment of a convicted offender would do harm to that person while yielding countervailing benefits to only one other. Since, from the perspective of the parties behind the veil, there is no reason to compromise the position of one person to benefit that of another, the parties would only authorize punishment where doing so could be shown to have an appreciable deterrent effect.²²¹

Eliminate the danger of wrongful conviction, however, and the parties *would* have a reason to punish where doing so would save only one potential victim from criminal violation: they would recognize that a culpable violation of the security and integrity of another seriously compromises that other's determinate conception of the good and her ability to pursue it. And although the parties would recognize that they themselves could turn out to be guilty offenders, doing harm to others in this way, they would nonetheless still recognize that the commission of serious offenses is wrong and creates unnecessary harms to innocent others. They would thus, in effect, condemn *themselves* as blameworthy and deserving of punishment

221. It might be thought that the parties, who could turn out to be either crime victims or targeted offenders, would be indifferent as to whether such punishment would be imposed under these circumstances. But this notion requires the assumption that, where the imposition on one's security and integrity is equally severe, the parties would view the harm of suffering punishment at the hands of the state despite being innocent of any crime as the equivalent of being victimized by crime. There is, however, arguably an even greater harm that comes to one being wrongfully targeted for punishment by the awesome coercive machinery of the state, with the broad social condemnation such targeting carries and the deep sense of violation that would come from experiencing punishment at the hands of official institutions supposedly put in place for the benefit and protection of citizens. For this reason, and because the authorization of punishment by the parties represents affirmative approval of the state's violation of the security and integrity of the target (whereas victims of crime generally receive only social approval and sympathy), I would expect that absent an appreciable deterrent effect, the parties would affirmatively refuse to authorize state punishment under the full conditions of partial compliance in cases where the balance in terms of the security and integrity of the crime victim and the target of punishment would otherwise be even.

should they turn out to be citizens who harm others in this way.²²² For this reason, the parties would accept that the guilty should be punished if it could be shown that, as a result, even one law-abiding citizen would escape the harm caused by the punished offense.

3. Disproportionately Severe Punishment of Serious Offenses

Are there any circumstances in which the parties deliberating for a partially compliant society would authorize punishments for serious offenders of *greater* severity than that suffered by their victims? The reasoning we have employed until now suggests otherwise. For as long as the parties believe that they themselves could end up, once the veil is lifted, occupying the position of guilty offender, we can expect them to reject alternatives that would place members of this group in a position below the worst-off position created by other available alternatives.

Until now, however, we have been proceeding on the assumption that the deterrent effect of a particular punishment would apply only to repeated incidence of that particular offense. For this reason, we have been measuring the harm caused by the punishment imposed on a convicted offender against the harm caused to a victim by the offense of conviction. Yet what if punishing a less serious offense could be shown to deter, not (or not only) further incidence of that particular offense, but also the commission of more serious offenses? What if, for example, it could be shown that imposing a disproportionately severe

222. Although, generally speaking, judgments of moral worth are beyond the reach of the parties, who are concerned "solely by considerations relating to what furthers the determinate conceptions of the good of the persons they represent," PL, *supra* note 16, at 315, this is not the case when it comes to violations of what Rawls calls the "basic natural duties, those which forbid us to injure other persons in their life and limb, or to deprive them of their liberty and property" TJ, *supra* note 14, at 314. When the violation in question is of this character, the parties—with their capacity to understand the difference between right and wrong—would be able to render such normative judgments even within the original position. For further discussion on this point, see *infra* note 224.

sentence on one who has been convicted of simple assault would deter future *murders* to an extent that proportionate punishment would not? In such a case, it could be argued that the disproportionate sentence for the simple assault would not place the target of punishment in a position significantly below the worst-off position that would otherwise be created. For, viewed with this broader lens, were a disproportionately severe punishment *not* imposed for the simple assault, the worst-off position would be that occupied by the murder victim.

It would therefore seem that, in cases where a disproportionately severe punishment for the commission of a serious offense may be shown to have a deterrent effect on future crimes of greater severity than even the punishment imposed, the parties would, following the reasoning of maximin, authorize such punishment. And notice that, in such a case, the deterrent effect need not be appreciable. For assuming that the disproportionate punishment imposed were still less severe than the harm that would otherwise be suffered by the victim of the more serious offense thereby deterred, the targeted offender would still not occupy a position worse off than would otherwise exist were no disproportionate punishment imposed. Although the convicted offender—who, remember, may well be innocent—now occupies the worst-off position, in placing him there we have nevertheless maximized the worst-off position. That position is still better than it would have been had disproportionate punishment not been imposed, despite the possibility that it is now occupied by a wrongfully convicted innocent.

The parties would reach this conclusion whether or not they could assume that all punished offenders are guilty. For even if the parties feared that they themselves might be wrongfully convicted and sentenced to a punishment of disproportionate severity, they would also fear being victims of the still more harmful crime that could otherwise have been deterred. Applying maximin reasoning would thus lead the parties to affirm a principle authorizing

punishments of disproportionate severity even where the danger of wrongful conviction is present.²²³

There is, however, a further possibility which *would* yield different results depending on whether the danger of wrongful conviction remains: cases in which punishments of disproportionate severity could be shown to have a greater deterrent effect on repeated incidence of the offense of conviction than would punishments of proportionate severity. In such a case, a disproportionately severe punishment would ensure that fewer members of society would occupy the disadvantaged position of the victim of the offense of conviction. At the same time, however, such a punishment would place the convicted offender—who now faces a punishment imposing a harm of disproportionate severity to that suffered by the victim of the offense of conviction—in a worse worst-off position than would otherwise have been created.

Assuming the full conditions of partial compliance, the parties would reject such a possibility. For where the parties recognize that they themselves could turn out to be wrongfully convicted and thus targets of punishment through no fault of their own, they would not authorize the state to punish in a way that could leave them in a worse worst-off position than would have otherwise existed. Were the danger of wrongful conviction eliminated, however, the parties would not necessarily reach the same conclusion. For arguably, where the parties could be certain that all those convicted of a serious offense were guilty as charged, the parties' sense that the act was wrong and the offender deserved punishment might be sufficient to ground an acceptance that the state may impose disproportionate

223. Arguably, the parties would go still further than this discussion suggests: under certain circumstances, the parties, following maximin, would authorize punishment of a severity equal to that of the more serious crime that punishment would deter. Here, however, a difference would emerge depending on whether the danger of wrongful conviction exists. Where it does, the parties would require that the deterrent effect be appreciable. For the reasons discussed in the text, the parties would not impose such a requirement where all convicted offenders are known to be guilty.

punishments on convicted offenders to protect innocents from suffering the harm of a serious offense.²²⁴

The issue here is thus whether the parties' sense of the wrongfulness of the act would lead them to accept a deviation from maximin that would place those guilty of serious offenses in a worse worst-off position than would otherwise exist. If so—and recall that, if so, this approach would be undertaken only where any danger of wrongful conviction had been eliminated—the question then becomes whether there would be any degree of disproportion too great for the parties to accept. What if legislators concluded that incarcerating convicted muggers for, say, thirty years could be expected to yield a greater reduction in the number of overall muggings than would a punishment proportionate to the harm caused by a mugging?²²⁵ Would the parties' recognition of the violation

224. To suggest that the parties in the original position would consider citizens' relative moral desert might seem odd, particularly given Rawls's well-known view that assertions of moral worth are out of place behind the veil. See TJ, *supra* note 14, at 310-11 ("There is a tendency for common sense to suppose that income and wealth, and the good things in life generally, should be distributed according to moral desert. . . . Such a principle would not be chosen in the original position."). For, as Rawls understands the quality of moral worth, it is that quality which leads citizens to want "to act in accordance with the principles [of justice] that would be chosen in the original position." *Id.* at 312. Citizens on this view are thus unable even to manifest their moral worth until after the principles are chosen and the veil has been lifted. And because determination of the bases on which the distribution of society's primary goods is to be effected is necessarily prior to demonstrations of moral worth, judgments about the relative moral worth of various citizens could not possibly enter into the calculus as to the size of individuals' respective distributive shares.

Rawls's view in this regard, however, is particular to the conditions of the well-ordered society. Where crime is a social reality, in contrast, it is possible to identify the morally unworthy prior to the selection of the principles of punishment in this narrow sense: they are those who commit serious offenses against others. For as Rawls himself notes, the prohibition on what we have here been calling "serious offenses" is in place to "uphold basic natural duties, those which forbid us to injure other persons in their life and limb, or to deprive them of their liberty and property . . ." *Id.* at 314. Thus although the parties behind the veil cannot yet know who in particular will commit these wrongs, they are nevertheless still able to recognize—and to condemn—the actions that constitute wrongs and thus the nature (in this respect, at least) of the morally undeserving, in advance of any other agreement. For helpful discussion on this issue, see Scheffler, *supra* note 34, at 978.

225. I am assuming here that the imposition on the security and integrity of a

done by a mugger to the security and integrity of his victim lead them to accept such a severe departure from maximin? Or, where no danger of wrongful conviction exists, would a rough proportionality principle instead be among the principles of punishment the parties would affirm behind the veil?

The answer to this question is not clear. On the one hand, the parties would recognize the wrongfulness of committing serious criminal offenses, and thus might not view as unjustifiable the imposition of punishment on guilty offenders—even should they themselves wind up among the offenders facing punishment. This sense of the fittingness of punishment in such cases might well be sufficient for the parties to accept even punishments of greatly disproportionate severity where their imposition would protect a law-abiding person from suffering the violation of a serious offense of any description. On the other hand, the same sense of justice that would lead the parties to condemn serious criminal offenders might also lead the parties to recognize that some offenses are more blameworthy than others, and to conclude that, although one who is guilty of a serious offense should expect and would rightly receive some punishment for his or her offense when doing so would improve the position of an innocent citizen vis-à-vis their security and integrity, there ought to be some limits.

Admittedly, viewing the question through the lens of the parties' sense of the wrongfulness of serious offenses takes us outside a maximin analysis. But even to consider the possibility of disproportionately severe punishment where such punishment would serve to deter the

convicted mugger of thirty years spent in prison is considerably greater than that imposed on the victim of his mugging, who would no doubt have suffered considerable distress, trauma, and anxiety for a prolonged period, but would be expected to recover well before thirty years had elapsed. Even if the victim were particularly sensitive and never completely lost his feelings of trauma, he could nonetheless, after some time had passed, be expected to get on with building his life and pursuing his preferred ends well within that time. For the offender, such possibilities would be largely precluded for the duration of his thirty-year sentence.

commission of crimes imposing harms of less severity is *already* to go outside maximin to consider placing the offender in a worse worst-off position than otherwise existed. And if, as I have suggested, it is the recognition that the offender has done a wrong and thereby affirmatively compromised the security and integrity of the victim that might justify this deviation from maximin on the part of the parties, it is important to recognize the way this same concern might serve as a limiting principle on the extent of any such deviation.

4. Disproportionately Severe Punishment of Non-Serious Offenses

Until now, we have been considering the extent to which the parties would authorize the disproportionate punishment of serious offenders. What, however, of the case in which punishing one convicted of a *non-serious* offense could be shown to have an appreciable deterrent effect against a serious offense?²²⁶ As we saw above, the parties assuming the full conditions of partial compliance would reject a principle authorizing punishment where doing so would deter the further commission of non-serious offenses. For doing so would allow the violation through incarceration of the security and integrity of the punished offender without any countervailing benefit in this regard for potential victims. However, where punishment for a non-serious offense could be shown to have an appreciable deterrent effect against serious offenses, the parties *would* arguably authorize such punishment, so long as the severity of the punishment was equal to or less than that of the serious offense thereby deterred. They would do so on

226. Note that in the case of punishing a non-serious offense, *any* sentence imposed is necessarily disproportionate to the offense, since non-serious offenses as we have defined them do not violate the victim's security and integrity. However, incarceration on the order we have been considering it here—i.e., of a length sufficient to represent a serious compromising of the security and integrity of the incarcerated—*would* represent such a violation. See *supra* p. 387 and note 207.

the same terms, and for the same maximin reasoning, on which they would authorize disproportionately severe punishment for serious offenses: doing so would maximize their prospects, even recognizing that they might turn out to be non-serious offenders facing disproportionate punishment. For they would also recognize that they might turn out to be among those citizens—necessarily greater in number than the targeted offenders, under the standard of the appreciable deterrent effect—spared from the serious offenses thereby deterred.

Notice that, in this case, the parties would again authorize punishment on precisely the same terms whether or not the danger of wrongful conviction exists. For here, even offenders guilty of non-serious offenses have done nothing to violate the security and integrity of their fellows. And where the targeted offender has done nothing wrong in this regard, the parties' sense of justice would recognize the entitlement of both guilty offenders and the wrongfully convicted to be protected from a violation of *their* security and integrity. Thus on both scenarios the parties would require that punishment only be imposed on non-serious offenses if it could be shown that doing so would have an *appreciable* deterrent effect—that is, that punishment would be reasonably certain to deter at least two future serious crimes for every offender punished²²⁷—and where the punishment imposed would be less severe than the harm averted in this way.

5. The Parsimony Principle

Although the parties under certain circumstances would endorse punishments even of severity disproportionate to the offense of conviction where necessary to achieve either an appreciable deterrent effect (under the full conditions of partial compliance) or any deterrent effect (where the danger of wrongful conviction has been

227. Otherwise, as we have seen, the principle would in effect allow the victim and the offender to switch places in terms of their welfare, where no sufficient reason would exist to justify disadvantaging the offender.

eliminated²²⁸), it is important to recognize that they would nevertheless, all things being equal, prefer still more a punishment imposing a hardship on the convicted offender less severe than that imposed on the victim by the offense. For so long as the deterrent effect is the same, the parties would, following *leximin*, prefer a situation that improves the position of the worst-off person who stands to be affected to one that does not. And where additional punishment would yield no deterrent effect, it would for this reason be rejected by the parties.

This attitude of parsimony is a key feature of the principles of punishment the parties would adopt, whether or not there is a danger of wrongful conviction. For in both cases, punishment imposed where it would have no (or no further) deterrent effect would place the offender in a worse-off position than she would otherwise occupy all other things being equal. It would thus be merely gratuitous, serving no legitimate purpose. The parties would therefore reject any principle on which punishment under these circumstances would be permitted, and would instead, consistent with *maximin*, adopt what we can think of as the "parsimony principle,"²²⁹ on which in all cases punishment must be no more severe than necessary to achieve the relevant deterrent effect.

The parsimony principle would apply, moreover, even were there no danger of wrongful conviction, for where punishment has no deterrent effect, the parties would have no interest in such punishment. This is so even viewing the question from the perspective of potential crime victims, for where punishment would violate the parsimony principle, any further punishment would by definition have no relevant benefit for the occupants of this position. Thus, the parties would reject gratuitous punishment even of the guilty.

228. With the one exception regarding the punishment of non-serious offenses; see *supra* pp. 399-400.

229. This imperative of parsimony is also an important feature of the theory of punishment developed by Braithwaite and Pettit. See Braithwaite & Pettit, *supra* note 39.

6. Determining Evaluative Standards

Plainly, a satisfactory method for assessing the relative severity of crimes and punishments and for evaluating the deterrent potential of alternative sentencing schemes is crucial to the effective and consistent implementation of the principles thus far identified. The difficulty, however, is that there is likely to be much disagreement over such methods and the conclusions reached by those who apply them. Thus the selection of standards by which such determinations are made becomes itself an important element—ultimately perhaps, the most important element—in ensuring legitimate punishment. As we will see, at the legislative stage, when the principles of punishment are translated into actual policies, even people reasoning in good faith may be expected to disagree and to reach different conclusions.²³⁰ It is still, however, possible at this initial stage for the parties to agree upon standards that will narrow the range of anticipated disagreement and maximize the likelihood that the conclusions drawn in any particular case at the legislative stage are consistent with the demands of the principles, including the parsimony principle, which the parties would select in the original position.

The parties may be expected to insist on two conditions for the determination of the relative severity of particular crimes and punishments and of the deterrent effect of a particular punishment scheme. First, we can expect the parties to require that such analysis be grounded in evidence and methods of reasoning which Rawls has come to associate with “public reason,” that is, “presently accepted general beliefs and forms of reasoning found in

230. This inevitable disagreement will have profound implications for the possibility of certainty as to whether the punishments actually imposed are consistent with the principles of legitimate punishment. It is for this reason, Rawls concludes, that “[o]ften the best that we can say of a law or policy is that it is at least not clearly unjust.” TJ, *supra* note 14, at 199. It is also for this reason that Rawls dispenses at the legislative stage with the assumption of unanimity he applies to the outcome of deliberations in the original position. See *id.* at 139. For further discussion of the legislative stage, see *infra* part V.B-C.

common sense, and the methods and conclusions of science when these are not controversial.”²³¹ For, as Rawls argues, adopting any other standard of reasoning “would involve a privileged place for the views of some over others, and a principle which permitted this could not be agreed to in the original position.”²³² And second, the parties may be expected to require a showing that the benefit in terms of citizens’ security and integrity claimed on behalf of a particular punishment scheme is not “merely possible or in certain cases even probable, but reasonably certain or imminent”²³³ For given the parties’ recognition that they themselves could end up on the receiving end of punishment, the parties would not agree to principles that compromised these goods on the basis of mere speculation or vague anticipated future benefits. If any punishment is to be legitimately imposed, those parties advocating its imposition must demonstrate convincingly, in terms that all could be expected to accept, that this imposition on the security and integrity of targets of punishment is immediately necessary or at least reasonably certain to result in greater protection for the law-abiding.²³⁴ Unless this burden can be met, the parties would conclude, no punishment may be legitimately imposed.

231. PL, *supra* note 16, at 224; see also TJ, *supra* note 14, at 213 (insisting that citizens may restrict liberty only when doing so accords with “evidence and ways of reasoning acceptable to all,” and “supported by ordinary observation and modes of thought (including the methods of rational scientific inquiry where these are not controversial) which are generally recognized as correct”). According to Rawls, such broadly acceptable standards are necessary if the conclusions reached are to comport with the liberal principle of legitimacy, on which citizens’ “exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational.” PL, *supra* note 16, at 217.

232. TJ, *supra* note 14, at 213.

233. *Id.*

234. By my use of the term “law-abiding” here and below, I do not intend implicitly to affirm disproportionately weak protection for citizens who might otherwise have committed a crime. Instead, I merely mean to distinguish the victims of crime, who are not in their victimhood criminally culpable, from the actors who commit the crimes against them.

In the remainder of this article, I refer to the foregoing conditions collectively as the requirement that asserted conclusions required by the principles be "reasonably certain."

C. The Inherent Illegitimacy of Wrongful Conviction

Assuming the full conditions of partial compliance, the parties would thus adopt a principle authorizing the imposition of punishment where doing so would achieve an appreciable deterrent effect against future offenses causing harm of equal or greater severity than the punishment imposed. This principle satisfies the terms of *leximin*. Yet, as the parties are aware, given the conditions of partial compliance, such a principle would necessarily mean that some number of innocent citizens will be wrongfully convicted and suffer serious harm at the hands of the state, despite having done no harm to anyone.

That this is so does not mean that the parties would accept the punishment of innocents as a just exercise of the state's power to punish. For there is a difference between a result that is the best that can be achieved under the circumstances, and a result that is widely understood and accepted as inherently just and fair. Although, given the non-ideal conditions of partial compliance, the parties would opt for principles that carry the risk that innocents will be wrongfully convicted and targeted for state punishment, they would nonetheless condemn as inherently illegitimate any particular instance of punishment where the target is in fact innocent of the crime charged. For anything else would authorize the state to actively and affirmatively compromise the security and integrity of an innocent member of society in precisely the way the parties would condemn were the agent of the harm a fellow citizen.

To see why this is so, consider the implications of a serious offense for the crime victim. Physical and psychological security are violated and even destroyed; life plans are derailed or negated altogether; bonds to family

and friends, to work, home, and community, are disrupted; future possibilities are diminished greatly, perhaps permanently. This is the affront of crime: that one person, to satisfy his own interests, takes it upon himself to intrude so violently and dramatically into the life and well-being of another who has done nothing to merit such treatment. For this reason, the parties would not only fear crime, but would also deeply resent the agent of any such violation. They would thus not only welcome the deterrent effect of any punishment imposed, but would also view it as to some degree deserved.²³⁵ And this is so despite the parties' recognition that they themselves could turn out to be guilty offenders—for in that event, the parties would understand that it is they themselves who have done wrong and thus earned the rightful resentment of their fellows. As guilty punished offenders, therefore, they would continue to fear the effect of punishment, but they would not in turn resent the punishment imposed on them.²³⁶

In cases where the convicted offender is in fact innocent, however, her experience is analogous, not to that of guilty targets of punishment, but to that of the crime victims themselves. For like crime victims, the wrongfully convicted offender has done nothing for which she may fairly be resented or for which the punishment imposed could be viewed as deserved. To the contrary, in such a case, it is the wrongfully convicted offender who would have cause to resent the intrusion on her security and integrity punishment represents. For here, although she

235. See Scheffler *supra* note 34, at 969 (“[R]esentment without an ostensible desert basis is not resentment.”) (quoting Joel Feinberg, *Justice and Personal Desert*, in *Doing and Deserving* 71 (Joel Feinberg, ed. 1970)). See also *supra* note 224.

236. At least, they would not resent it where the severity of the punishment bore some relationship to the harm caused by the offense of conviction. Where the punishment is greatly disproportionate to the severity of the offense, it seems quite possible that the parties *would* resent such a prospect, even recognizing the wrong of the harm caused by the offense itself. It is this sense that lies behind the suggestion, raised above, see *supra* pp. 397-98, that the parties might endorse a rough proportionality principle—or at least forbid punishments of severe disproportionality—even when the convicted offender is known to be guilty as charged.

has committed no wrong, it is *her* physical and psychological security that have been violated; *her* life plans that have been derailed; *her* ties to family, friends, and community that have been disrupted; *her* future possibilities that have been compromised arbitrarily. In short, imposing state punishment on such a one effects the same harm—the same sudden, arbitrary, and violent intrusion into the life of an innocent person—that violent crimes cause their victims, and as with crime, it uses the target to promote the interests and purposes of others. The same sense of justice that would lead the parties to condemn the commission of serious offenses as morally blameworthy would thus necessarily lead the parties to condemn the punishment of innocents as equally wrongful.²³⁷

This is not to say that the parties would, on this basis, reject any principle that could lead to wrongful conviction. Under conditions of partial compliance, it would be impossible for the parties to vindicate their sense of the inherent illegitimacy of wrongful conviction unless they rejected any state punishment at all. And for reasons we have seen, the parties—who are most centrally concerned with protecting their security and integrity to the greatest extent possible—would prefer a principle that carries with it some danger of wrongful conviction to a principle allowing no punishment, and thus no deterrence, at all.²³⁸

237. For this reason, I disagree with Pogge, who suggests that under certain circumstances the parties would have reason to affirm wrongful conviction as an appropriate option. See Pogge, *supra* note 34, at 259. See also *supra* note 224 (explaining the sense in which the parties are able to judge the blameworthiness of particular criminal actions behind the veil).

238. This is why the parties would not select a principle vindicating any explicit preference—however mild—for a retributive theory of punishment. Depending on how it is crafted, such a principle might comport with the parties' sense of who deserves to be punished, but it could also easily lead to the imposition of gratuitous punishment—punishment that achieves no benefit in terms of protecting anyone's security and integrity—while diminishing, perhaps greatly, the security and integrity of the target. And the parties, knowing that they could end up the targets of state punishment as well as the victims of crime, would reject such an outcome.

At the same time, the parties' condemnation of wrongful conviction and their recognition of the blameworthiness of guilty offenders overlap to a great

At the same time, however, the parties would continue to insist on the inherent illegitimacy of wrongful conviction. This insistence would mean that, in addition to adopting a principle authorizing punishment where doing so would yield an appreciable deterrent effect, the parties would also adopt a further principle, one that requires the state to do all it can to reform the system, both to reduce as much as possible the incidence of wrongful conviction and to ensure mechanisms for detecting and remedying the mistakes that have already occurred in this regard.²³⁹ For the wrongful conviction of innocents, while inevitable, would never cease to be an affront, a source of resentment. While leaving unchanged the maximin calculus explored above, this principle would address this affront and mitigate the effects of wrongful conviction on the strains of commitment.

Notice that in seeking to prevent the wrongful conviction of innocents, the parties lose nothing in terms of the potential deterrent benefit of punishment. If anything, they are bolstering this deterrent potential. For a system that did nothing to prevent or redress wrongful conviction could have, if anything, a limited deterrent effect; if citizens are sufficiently unsure that those individuals who are being punished are actually guilty as charged, would-be offenders would no longer perceive the necessary connection between wrongful conduct and punishment necessary to make deterrence work. Nor is the answer that to preserve a deterrent effect the state should keep secret the fact or possibility of wrongful conviction. For not only

extent with the motivating views of retributivists. The difference is that, on the theory I develop here, the moral desert of the offender is not a sufficient condition for imposing punishment or for setting the intensity of the punishments imposed—although it is a necessary condition for viewing any such punishment as legitimately imposed.

239. The possibility of mistakes would also affect the calculus of the deterrent effect. We have assumed that, notwithstanding the danger of wrongful conviction, the system is sufficiently trustworthy to make some deterrent effect possible. But to the extent that the system continues to be perceived as untrustworthy to any extent, those calculating the deterrent effect of any punishments will need to make some allowances based on this perception. This, however, is largely a matter for the legislative stage.

does Rawls stipulate the publicity condition²⁴⁰—with good reason—as a basic feature of principles of justice in a liberal democracy, but in addition, given what the parties know about the state's tendency under conditions of partial compliance to abuse its power in ways that can seriously violate the security and integrity of ordinary citizens, the parties would not trust state officials sufficiently to be willing to authorize such secret action.²⁴¹ And with the publicity condition in place, the punishment of innocents would achieve no benefit that the parties might otherwise want to capture.

D. Enumerating the Principles

It is now possible to identify a number of principles to which the parties, following maximin and assuming the full conditions of partial compliance,²⁴² may be expected to agree:

1. There shall be no incarceration for non-serious offenses, unless doing so would appreciably deter²⁴³ the commission of serious offenses.²⁴⁴
2. Punishments of incarceration, when imposed for serious offenses, may be only as severe as necessary to

240. See TJ, *supra* note 14, at 133, 177; PL, *supra* note 16, at 317.

241. See Shklar, *supra* note 3.

242. As we have seen, there would be some divergence between the content of the parties' agreement on the full conditions of partial compliance and that which would be reached where the danger of wrongful conviction were eliminated. Because I am primarily concerned with the principles that would govern in the scenario closest to our own, I express the principles here in terms of the full conditions of partial compliance.

243. For simplicity's sake, in the formulation of these principles I use the term "appreciably deter" to describe punishments necessary to achieve a reasonably certain or imminent appreciable deterrent effect.

244. This principle is a variant on the parsimony principle. I articulate it separately to emphasize that within the class of offenses that qualify as non-serious, the parties would reject as illegitimate any incarceration of offenders. Note that this principle would not rule out the establishment and imposition of less serious sanctions for the commission of those offenses judged non-serious.

appreciably deter offenses causing harm of equal or greater severity (the parsimony principle).

3. Before any punishment may be imposed, its deterrent effect must be shown to be reasonably certain or imminent, on the basis of standards and modes of reasoning acceptable to all.
4. Consistent with these principles, the state must do all it can to reform the criminal justice system in order to reduce as much as possible the danger of convicting the innocent or retaining them in custody.²⁴⁵

A central premise of my argument has been that the exercise of state power in a liberal democracy is legitimate when it is consistent with principles that parties under fair deliberative conditions would agree are just and fair. If this is indeed so, and if I am right that these are the principles to which the parties would agree under such conditions, we are now in a position to recognize the character of legitimate punishment in liberal democracy: state punishment that is consistent with the principles here articulated.

E. Incarceration under Inhumane Conditions

1. The Problem of Inhumane Punishment

Until now, we have assumed that the only available form of punishment is incarceration for a specified time under humane conditions.²⁴⁶ But what if the parties also

245. As long as there is crime, the parties will view punishment as a necessary evil. For this reason—and because the parties hope to avoid crime as much as they hope to avoid punishment—the parties deliberating behind the veil would take a further position, which because of its character may best be thought of not as a principle but as a background condition. That is, the parties would agree that the state must take whatever means are available to reduce the incidence of crime, in order both to protect citizens from that source of harm and to reduce the overall amount of punishment that the state could impose consistent with the principles.

246. When I speak of humane conditions of confinement, I mean conditions that are safe, healthy, and as comfortable as possible under the circumstances.

had available to them the option of punishment under other, less humane conditions—conditions that degrade, humiliate, or otherwise seriously compromise or undermine altogether the central features of the target's moral personhood?²⁴⁷ Call punishments imposed under such conditions inhumane punishments.²⁴⁸

What is it that makes inhumane punishment a special problem for the parties? Imposing such punishment would represent a violation of a different order than even the violations we have thus far been contemplating. For, as we have seen, it is our capacity to develop and pursue a meaningful conception of the good and our capacity to understand and act from a sense of justice that make us moral persons. Absent these distinctive features of moral personhood, we would be unable to formulate our own conception of the good, much less to pursue it.²⁴⁹ For this reason, the parties may be said to have a highest-order interest—more urgent than even the protection of their security and integrity as we have conceived it—in not

247. As above, I leave the task of determining precisely what conditions would fall into this category to policymakers at the legislative stage.

248. Given the considerable restrictions on freedom of action that attends incarceration in any form, it might be thought that, on this definition, incarceration would per se qualify as inhumane. But what is at stake when inhumane punishment is imposed is not the obstacles it creates to the realization of one's aims and ends. It is rather its assault on the very capacities of the targets to recognize their own interests and know how to pursue them. Under humane conditions of incarceration, it is still possible for prisoners to form bonds with others, maintain contact with family and friends outside the prison walls, and identify and pursue other interests, albeit in a highly constrained way. Under inhumane conditions—for example, those of being forced to live for extended periods in fear of physical abuse or sexual assault—the subject of such punishment would lack even the capacity to understand himself as a subject of interests worth fulfilling—beyond, that is, the basic interest of survival. It is this particular form of violation, and not the more general (although serious) compromise of one's security and integrity, that makes punishment inhumane.

249. See PL, *supra* note 16, at 75-76 (describing the primary goods as the "social background conditions and general all-purpose means normally needed for developing and exercising the two moral powers and for effectively pursuing conceptions of the good with widely different contents"); see also *id.* at 315-24 (describing the relationship between the development and exercise of one's capacity for a sense of justice and the possibility of developing and furthering one's determinate conception of the good).

having the bases of their moral personhood seriously compromised or undermined altogether.

Given this highest-order interest, the first principle the parties would adopt when contemplating the possibility of inhumane punishment would be one denying the state the authority to impose inhumane punishments where doing so would be gratuitous, not required to improve the position of the worst-off person vis-à-vis their highest-order interest in avoiding inhumane treatment. Otherwise, the target of such punishment would be placed in a worse worst-off position vis-à-vis this interest without any countervailing improvement in the highest-order interests of anyone else. That is, the parties would agree to a fifth principle, that:

5. Punishment may not be gratuitously inhumane.²⁵⁰

This requirement has two main implications. First, where the desired deterrent effect may be achieved with either humane or inhumane forms of punishment, the state is obliged to impose the former even if there may be other reasons—reasons of policy, perhaps²⁵¹—why the latter may be preferred. And second, where it is determined that inhumane punishment would be necessary to achieve the relevant deterrent effect (and where doing so would be consistent with the terms on which the parties would authorize the imposition of such punishments²⁵²), the inhumane punishments imposed may not be, in either duration or form, more severe than necessary to achieve the relevant deterrent effect. For the imposition of any

250. This principle is implicit in the principles already identified. I articulate it separately to underscore the particular urgency the parties would feel that the state not exceed this authority when the very moral personhood of the target is at stake.

251. Given the priority of the parties' higher-order interests, the parties would not accept disadvantages in terms of these interests, even if they were compensated with advantages in terms of their lower-order interests. This priority would be even greater for the highest-order interests we have just identified, and thus the possibility of effecting such a compromise would be even more unappealing to the parties in this case.

252. For a discussion of these terms, see *infra* part IV.E.2.

such punishment beyond this point would be merely gratuitous, violating the highest-order interests of the target while providing no benefits to anyone else in terms of their highest-order interests.

2. Inhumane Punishment and the Choice among Evils

What, however, of circumstances in which the imposition of inhumane punishment would not be gratuitous, but would instead be consistent with the requirements of maximin? Are there such cases? The parties, as we have just seen, have a highest-order interest in protecting the central capacities comprising their moral personhood. And in a partially compliant society, it is not only state punishment that presents a threat to this interest, but also the commission of crimes; some crimes are so brutal that they too would qualify as inhumane, compromising or undermining altogether the moral powers of the victim. Given this fact, and given the parties' concern with preserving their highest-order interests, it might be thought that, consistent with maximin, the parties would without hesitation authorize the state to impose inhumane punishment where doing so is reasonably certain to appreciably deter offenses of a severity equal to or greater than the punishment imposed.

There may, however, be reason to think that, under the full conditions of partial compliance, the parties reasoning according to maximin would reach a different conclusion regarding the imposition of inhumane punishments. To see why this is so, consider the implications of such punishment for the target. Authorizing the state to impose inhumane punishment on a convicted offender is to authorize the state actively and affirmatively to strip away the essential components of the target's moral personhood. Such action would, by design, seriously compromise or perhaps undermine totally the capacity of the target to develop and pursue any kind of meaningful existence, however circumscribed. It would, moreover, exclude the target from the circle of moral

persons, placing the ongoing preservation of the conditions of his moral agency beyond the realm of social concern. And now consider that, under the full conditions of partial compliance, the convicted offender subject to this treatment may well be innocent of the crime charged. Above, I argued that, notwithstanding the parties' ongoing recognition of the inherent illegitimacy of wrongful conviction and the possibility that some innocents would be subject to state punishment, they would still authorize the state to punish convicted offenders under certain circumstances. But notice that when the punishment in question is incarceration under humane conditions, the parties would still know that, however things turned out for them, they would at least maintain the essential components of their moral personhood and occupy a place within the moral boundaries of society. In the case of inhumane punishment, however, the parties must consider the possibility that they could turn out, for morally arbitrary reasons, to be stripped of the bases of their moral personhood and excluded from the realm of social concern. For this reason, where the punishment in question is inhumane, the parties may well refuse to authorize such punishment where the danger of wrongful conviction exists, finding the strains of commitment created by such an agreement to be too great to bear.

There is, of course, arguably a flaw in this reasoning: were the parties *not* to authorize such punishment where doing so would be reasonably certain to appreciably deter crimes of a severity equal to or greater than the punishment imposed, they would thereby create the conditions under which a greater number of citizens would come to suffer inhumane treatment at the hands of criminal offenders than would have been the case had the inhumane punishment in fact been imposed. And, it might be thought, however strongly the parties would resist the possibility that, for morally arbitrary reasons, they could wind up targets of inhumane punishment at the hands of the state, they would resist equally strongly the possibility of winding up the victims of inhumane crimes. For this

reason, the parties following maximin could be expected to authorize the imposition of inhumane punishment despite the ongoing danger of wrongful conviction.

Admittedly, were the experience of a crime victim suffering inhumane treatment at the hands of a criminal and that of a target suffering inhumane punishment at the hands of the state equivalent in character, both the logic of maximin and the imperative of the strains of commitment would lead the parties to authorize inhumane punishment under the constrained conditions already articulated. No doubt there are arguments to support this view. Here, however, I want to suggest that there may be a salient difference that makes subjection to inhumane punishment by the state an even greater violation of one's moral personhood—and thus an even greater horror—than that suffered by victims of inhumane crime. For when convicted offenders are incarcerated under inhumane conditions, they are, as already mentioned, stripped of their membership in the human community, placed beyond the circle of moral concern. This condition is more than a form of moral exile (although it is that). Worse still, the targets of such punishment become less than human in the eyes of society, not only demeaned and degraded but seen to be deserving of such treatment. Crime victims, in contrast, are never viewed in this light. Their degradation is never affirmed as deserved. To the contrary, it is precisely because maintaining the basis of the victim's moral personhood is of deep and ongoing concern to the collective that such crimes—and their perpetrators—are so roundly condemned.

If the foregoing captures a meaningful difference between these two (horrible) experiences, it follows that, even on the terms of maximin, the parties would refuse to authorize inhumane punishment where the danger of wrongful conviction exists.²⁵³ For to do so would be to

253. Here, the strains of commitment arguably bolster this conclusion, for the parties, choosing as they are the principles that will establish the terms of state punishment for all time, could not enter an agreement on the terms of which they could be placed outside the circle of social concern, their moral personhood denied

create a worse worst-off position than would otherwise exist, notwithstanding the lost deterrent effect on crimes of lesser—although still great—severity. This difference may also be articulated in terms of the primary good of self-respect, which Rawls suggests may be the most important primary good. As he argues, self-respect—which provides a “secure sense of our own value”—is not merely an internal quality of individuals, but “depends upon and is encouraged by certain public features of basic social institutions,” including “how people who accept [social] arrangements are expected to (and normally do) regard and treat one another.”²⁵⁴ If the state’s affirmative denial of one’s status as a moral person is the ultimate undermining of one’s self-respect, then it follows that being the deliberate target of inhumane punishment by the state would put one in a worse worst-off position than would otherwise exist, even in cases where the imposition of inhumane punishment would have yielded a reasonably certain deterrent effect against inhumane crimes.²⁵⁵

However, even if I am right that authorizing such punishment would constitute a departure from maximin, the parties might well not have the same resistance to the imposition of inhumane punishment were the danger of wrongful conviction eliminated.²⁵⁶ For notwithstanding their recognition that they could turn out to be guilty

and even negated, despite their having done no harm to anyone.

254. PL, *supra* note 16, at 319.

255. It might of course be argued that inaction on the part of the state where preventative action was possible constitutes an effective denial, akin to that suffered by the targets of inhumane punishment, of the moral personhood of those citizens who wind up victims of inhumane crimes due to the state’s failure to deter their assailants. For the reasons I offer above, I do not believe this to be the case. But if I am wrong in this regard—if, that is, the parties would take it that the harm is equal as between victims of inhumane crimes and targets of inhumane punishment—then notwithstanding the possible innocence of the target, the parties *would* authorize inhumane punishment where doing so would be reasonably certain appreciably to deter the future commission of inhumane offenses.

256. The reasoning in this case would be analogous to that employed in considering the possibility of the parties’ authorizing disproportionately severe punishment where such punishment would appreciably deter harms less severe than the punishment.

offenders, the parties would recognize the wrongfulness of criminal action and the blameworthiness of the offender who commits them. Thus, where the guilt of the offender is certain and the imposition of disproportionately severe inhumane punishment—with the contempt and exclusion from the circle of moral concern such punishments necessarily entail—would mean that some number of innocents would thereby avoid being victims of inhumane crimes, the parties might well authorize such a departure from maximin. And that is so notwithstanding that they could thereby wind up the targets of inhumane punishment.

Or at least, they might do so where, in addition to the stipulations already noted, the offense of conviction is itself not merely a violation of the victim's security and integrity but is also inhumane as we have understood this term here. Where convicted offenders are themselves guilty of inhumane crimes, they have indicated their indifference—even their contempt—for the moral personhood of another. The parties would therefore understand that if it is they themselves who commit such grave harms against others, they would have thus earned from their fellow citizens rightful resentment, severe contempt, and an unconcern with the preservation of essential components of their moral personhood. Where, however, the guilty offender has not acted in this way—where, that is, the offense of conviction is serious but not inhumane—the same cannot be said. For the parties' possible willingness to depart from maximin reasoning under these conditions arguably hinges on the recognition by the parties that the convicted offender has committed a blameworthy act and thus in some sense deserves the punishment imposed.²⁵⁷ Where the punishment is greatly disproportionate to the severity of the offense—as it would be were the punishment imposed inhumane while the offense of conviction is serious but not inhumane—it seems quite possible that the parties

257. See *supra* note 224 for a discussion of the parties' capacity to assess the desert of criminal offenders.

would resent such treatment, notwithstanding their recognition of the wrong of the harm done by the serious offender.

The question, then, is whether the parties would view the blameworthiness attached to the commission of a serious but not inhumane offense as justifying their potential exclusion from the circle of moral citizens. If not, the measure of disproportionate severity that inhumane punishment would represent would lead them to reject the departure from maximin reasoning the imposition of such punishment would require.²⁵⁸ Certainly, so long as the danger of wrongful conviction exists, the question would not arise in any case. It is, however, to be expected that, even under the full conditions of partial compliance, the guilt of some offenders charged with inhumane crimes will be incontrovertible. And in such cases, for the reasons just articulated, the parties may well under some circumstances authorize the imposition of inhumane punishment. It is thus possible that, on the framework I have adopted, inhumane punishment could be construed as legitimate under some circumstances.

There is a real danger that such a conclusion, however carefully qualified and circumscribed, will be misconstrued, so let me be clear. I understand this possibility to be a very narrow one indeed. I have here gone only so far as to say that such punishment could be legitimate only where

1. the imposition of inhumane punishment is necessary to deter offenses of equal or greater severity (and thus is not gratuitous);
2. the guilt of the offender is incontrovertible;
3. the offense of conviction is itself inhumane;
4. the deterrent effect can be shown to be reasonably certain or imminent; and

258. For recall that the authorization of such punishment by the state would constitute an affirmative and active negation by the state of the essential basis of the target's moral personhood, an affirmative denial of the humanity of the target by the society that crime victims—however inhumane the offense—never experience.

5. the punishment imposed is of a severity and duration no greater than necessary to achieve the anticipated deterrent effect.

And notice that, of these restrictions, the first, fourth, and fifth would obtain even if maximin would lead the parties to authorize inhumane punishment on terms consistent with the principles of punishment already identified.

Of these restrictions, I view the last as particularly significant. It says that punishments under inhumane conditions may not be of an indefinite or uncertain length. Rather, they must be calibrated as carefully as possible to ensure that they meet the conditions listed above. Wholesale incarceration under inhumane conditions, obviously out of bounds in the ordinary run of cases, would thus also constitute a violation of the principles beyond the point at which maintaining particular convicted offenders under inhumane conditions can be shown to be reasonably certain to achieve the relevant deterrent effect against future harms of equal or greater severity. This narrow grant of authority to punish with incarceration under inhumane conditions is thus no license for inhumanity. To the contrary, even assuming that the guilt of the offender is incontrovertible, the state's burden in these cases is a heavy one.

The central aim of this article has been to identify the principles of punishment that parties deliberating under suitably general conditions would accept as just and fair. We have now achieved this aim, identifying five principles of punishment the parties would adopt under the full conditions of partial compliance. We have, moreover, seen that by and large the same set of principles would also be chosen were the danger of wrongful conviction eliminated. Certainly, there are differences—most notably, perhaps, our late conclusion regarding the punishments of incarceration under inhumane conditions. But even here, the possibility is highly circumscribed, and in terms of the other principles, the differences are not that pronounced. There are two main differences. First, whereas on the full

conditions of partial compliance, the deterrent effect must be appreciable, where the danger of wrongful conviction has been eliminated, any deterrent effect will suffice. And second, where the disproportionately severe punishment of an offense would deter not a more serious offense, but only further incidence of the same offense, the parties under the full conditions of partial compliance would reject the imposition of disproportionate punishment. Absent the danger of wrongful conviction they would not.²⁵⁹ Still, the relative parity of our conclusions in this regard is perhaps the most surprising result to emerge from the above discussion.

My hope in pursuing this question has been to provide an understanding of the nature of legitimate punishment in a liberal democracy where, although crime is an inevitability, the punishment of criminal offenders also represents a considerable exercise of coercive state power. If, however, these principles are to provide a standard against which to evaluate the legitimacy of prevailing criminal justice policies, we must take a still further step: determining how these principles might be brought to bear on the problems of the real world. It is thus to this final consideration that we now turn.

V. FROM PRINCIPLES TO POLICIES

A. *Realizing the Ideal*

At present in the United States, there are over two million people behind bars—more incarcerated people per capita than in any other country.²⁶⁰ Granted, this fact alone tells us nothing about the legitimacy of the sentences being served. But the sheer scale of incarceration this number represents, particularly in light of the

259. Where, however, the parties know that all convicted offenders are guilty, they would arguably accept as just and fair a larger diminution in the security and integrity of the guilty to allow a smaller increase in the security and integrity of innocent victims.

260. See citations at *supra* notes 4-6.

disproportionate number of poor people and people of color in America's prisons and jails, signals the real possibility that the state is currently exercising a punitive power well beyond the bounds of legitimacy.²⁶¹ This possibility should trouble all members of American society, for it is in our name, and with our tacit assent, that our prisons are being filled.

Making sure that the imposition of state punishment in the United States is always fully consistent with the demands of legitimacy is an unattainable goal, but it should still be possible to shift prevailing criminal justice policies closer to the ideal of legitimate punishment. If we are to meet even this more modest goal, however, we must first determine how to transform the abstract principles we have here derived into workable criminal justice policies. Explaining this process is thus the first task of this part. I conclude that, although the details of any such policies must necessarily be left to legislators at the policymaking stage, the principles place significant substantive boundaries on the kinds of arguments legislators may legitimately offer, and on the policies they may ultimately endorse. This effect is achieved through a feature of what

261. Given the disproportion of poor people and people of color behind bars in the United States, one might think that any effort to establish the terms of legitimate punishment in liberal democracy would of necessity put issues of race and class front and center. The approach I adopt here, however, does no such thing; to the contrary, by locating the source of normative justifications for state punishment in the principles to which citizens would agree when deliberating behind a veil of ignorance, it self-consciously brackets issues of relative power, including the issues of race and class. Yet, perhaps paradoxically, it is only by bracketing these issues that we guarantee full and fair consideration of the interests and perspectives of society's least advantaged members. For when the powerful are free to exercise their power without constraint, they can draft sentencing policies safe in the knowledge that they themselves are highly unlikely to be subject to the punishments they authorize. They may therefore discount—and discount entirely, if they so choose—the interests of likely targets. If, however, one does not know the particulars of one's own identity, that luxury disappears, for under these conditions, the only way to protect one's own interests is to ensure the protection of the interests of all those members of society who stand to be affected by the particular measure under consideration. It thus is behind a veil of ignorance that the liberal ideal of moral equality may be most closely realized.

Rawls terms "the legislative stage," at which policy deliberations take place behind what we can think of as a "modified veil." This modified veil would allow legislators enough knowledge to fashion criminal justice policies suitable for the society for which they legislate. It would, however, continue to obscure the knowledge of legislators' personal particulars, which could skew the results in a way that serves the legislators themselves at the expense of other, less powerful members of society. Furthermore, as I suggest at the close of this part, although reasonable disagreement is inevitable on a range of questions likely to arise at the policymaking stage, we can nevertheless identify a number of criminal justice policies currently in force in the United States, the legitimacy of which may be called into question by a fair reading of the principles. If I am right in this regard, and if we are obliged, as members of a polity that is punishing to excess, to do what we can to restrain this illegitimate exercise of power, a reform of the policies identified in part V.D would be a good place to start.

B. The Modified Veil

In terms of the move from principles to policies, an immediate difficulty presents itself. That is, many questions that would need to be answered before the principles could yield specific and legitimate punishments in particular cases—which acts, for example, would qualify as serious offenses as we have defined them? what length of sentence for particular offenses would satisfy the demands of parsimony? how are we even to go about deciding these matters?—are as open to competing views and interpretations as any other contentious political questions. They are, moreover, as vulnerable to influence by individuals' knowledge of their own particulars and sense of their own interests as is the initial selection of the principles.

For this reason, it is not possible to look to the current political process to translate the principles of punishment

we have identified into criminal justice policies, and expect that the fruits of this process will thereby be legitimate. For in the legislative process as it currently exists, there is little to prevent legislators in the political arena from interpreting the principles or data or available options in ways that favor their own interests or those of their constituents at the expense of the interests of other, likely less fortunate and less powerful members of society.²⁶² To avoid this outcome, we need a process for translating the principles into policies, one that functions like the deliberative process of the original position. That is, we need a process that prevents the parties' knowledge of their own personal particulars from influencing the results and instead ensures consideration of the perspective of all citizens who stand to be affected by the policies ultimately selected.

The simplest solution to this problem would be to apply the methodology employed above to the task of translating our abstract principles into concrete policies. Unfortunately, the veil of ignorance as previously constructed would not do for this purpose. For the process of deriving workable criminal justice policies from abstract principles requires the parties to have more detailed knowledge of the society for which they are legislating than the full veil of ignorance allows. The parties behind the veil in the original position know simply "the general facts about human society."²⁶³ To craft criminal justice policies appropriate to a particular society, however, they need more than a general knowledge of the workings of human society. They need particularized knowledge: what sorts of

262. There are arguably aspects of the legislative process in the United States that seek to achieve this sort of broad consideration of interests: hearings, super-majority rules, some constitutional protections, etc. Yet each of these aspects is entirely consistent with legislators' continued commitment to the interests of their perceived constituents at the expense of the interests of less powerful members of society. Absent a clear commitment to adopting a measure of impartiality, legislators thus may—and will—adopt legislation without ever considering the implications for those most likely to bear the burdens thereby created.

263. TJ, *supra* note 14, at 137.

crimes threaten the citizenry of the society they are about to enter? To what sorts of harm do these crimes give rise? What sort of deterrent effect does the threat of incarceration for various periods have in this particular society? What available mechanisms are most effective for guarding against wrongful conviction? To answer these questions, the parties must know a good deal more about their particular social context than the veil as we have known it would allow.

In resolving this problem, I again follow the solution Rawls provides. Rawls proposes that, following the derivation of the principles, the parties enter what he calls the "legislative stage" in which they identify and enact into law the policies that best realize the principles previously selected.²⁶⁴ At this stage, although the parties continue to deliberate behind the veil, it is now thinner, allowing in the information about the particulars of their own society necessary if the parties are to make informed judgments, while at the same time still screening out the parties' knowledge of their attributes and personal particulars.²⁶⁵ In other words, although the participants would now have full access to all the specific facts of society necessary to ensure legitimate results, they are still required to deliberate as if they did not know anything about their own attributes or personal particulars.

In the context of working out actual criminal justice policies, this modified veil is of course just an ideal—it could not actually exist as a constraint on debate among actual people. But its acceptance as the appropriate deliberative standard against which to judge the positions

264. *Id.* at 198. This legislative stage is the third stage of what Rawls terms the "four-stage sequence." *Id.* at 195. See also *infra* pp. 424-25.

265. See *id.* at 200. As Rawls explains it,

[t]he flow of information is determined at each stage by what is required in order to apply these principles intelligently to the kind of question of justice at hand, while at the same time any knowledge that is likely to give rise to bias and distortion and to set men against one another is ruled out. The notion of the rational and impartial application of principles defines the kind of knowledge that is admissible.

Id. at 200.

taken by actual participants in the process would affirm the central importance of a stance of impartiality to the possibility of ensuring the legitimate exercise of state power. For when policymakers do *not* "step behind the veil" and thus accord due consideration to the interests of all members of society, there is a real danger that the policies adopted will instead privilege the interests of people who resemble themselves, and considerably discount the interests of those who do not. And given the dominance in legislative and policymaking circles of the most powerful members of society, such a failure will invariably result in policies that privilege the most powerful at the expense of other, less powerful citizens. Indeed, it is my sense that much of the illegitimate punishment currently being imposed in the United States stems from a failure of policymakers to take seriously the strong interests of the politically disenfranchised—in particular, the interests of poor people of color—in preserving their security and integrity and in retaining the capacities that underpin their moral personhood.

C. The Four-Stage Sequence and the Limits of the Legislative Stage

This legislative stage is in fact the third stage of what Rawls calls his "four-stage sequence" for deriving the principles and applying them to particular social contexts.²⁶⁶ The sequence begins with the first stage of the original position, in which the general principles are derived. The second stage, which Rawls terms the "constitutional convention," is that at which the parties "design a system for the constitutional powers of government and the basic rights of citizens."²⁶⁷ The constitutional principles agreed to at this second stage,²⁶⁸

266. See *id.* at 195-200.

267. *Id.* at 196-97.

268. Determining the particular constitutional design of political institutions and the political process the parties would choose at this stage is beyond the scope of this article. For simplicity's sake, I will assume that the parties would choose

together with the principles agreed to in the original position, would represent the terms against which any policies under consideration at the third legislative stage would be evaluated.²⁶⁹ In what follows, by my use of the term "principles," I mean to refer both to the principles agreed to in the original position and the principles agreed to at the constitutional stage.

Within the scope allowed by the principles, there is room for a number of different policy responses to any given problem. Thus, the framework established in the first two stages of the process cannot alone determine the content of the policies ultimately agreed to by the parties at the legislative stage. Certainly, this framework sets limits on the policies that would ultimately be chosen at this

some version of the structure of constitutional democracy that currently exists, a structure of separation of powers assuring checks and balances. Furthermore, because the parties would value above all the freedom to pursue their own conception of the good and to develop the moral powers that make possible the formulation of such a conception, I will, following Rawls, assume that the parties at this second stage would be sure to include in their constitutional scheme a commitment to the basic "liberties of equal citizenship": "liberty of conscience and freedom of thought, liberty of the person, and equal political rights." *Id.* at 197. Finally, we should expect at the constitutional convention stage an inclusion in the final agreement of the strongest possible procedural protections against convicting the innocent. Although the actual protections available may change as society evolves (a matter to be taken up at the legislative stage), at the very least, we can expect these protections to include the right to what has come to be known as due process; some sort of right against self-incrimination (to guard against false confessions made under threat of torture or other state-imposed pressures); and certainly a guaranteed and meaningful right to counsel.

269. At the fourth and final stage, the rules crafted at the legislative stage are applied "to particular cases by judges and administrators . . ." *Id.* at 199. Rawls also includes "the following of rules by citizens generally" as part of the workings of this final stage. *Id.* It is Rawls's position that no limits on self-knowledge are necessary at the final, adjudicative stage at which the policies and laws enacted by the legislature are to be applied. Yet any broad policies derived from the principles will necessarily remain at some level of abstraction, and will continue to require judgments and assessments of the available evidence if decisions are to be reached. Thus here too, it seems to me, decision makers will continue to be susceptible to the corrupting effects of the knowledge of their personal particulars that Rawls is so concerned to purge from the deliberations at prior stages. For this reason, I expect that some modified veil of ignorance, at least for the decision maker, would also be required at the last stage, in order to ensure that the policies chosen at the third stage remain as true in their implementation as the process of deriving the principles on which they were based.

stage; any policies that are plainly inconsistent with the principles would be inadmissible, ruled out from the start. But if, at the first stage, the general nature of the issue facing the parties and the abstract character of their available knowledge made possible an assumption of unanimity,²⁷⁰ the fact that reasonable differences of opinion exist as to what policies the principles require make widespread agreement over the policies to be enacted at the third stage unlikely.²⁷¹

There will thus inevitably be disagreement at the legislative stage over the content of the policies which best realize the abstract imperatives of the principles. Indeed, even with the most unassailable empirical evidence and analysis, legislators may reasonably draw different conclusions—as to, for example, the serious nature of particular offenses or the relative harm of particular sentences as compared with particular crimes. And to complicate matters even further, under non-ideal conditions,

270. See TJ, *supra* note 14, at 139:

[In the original position,] it is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. . . . If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached.

271. See *id.* at 198-99 (“[T]he question whether legislation is just or unjust, especially in connection with economic and social policies, is commonly subject to reasonable differences of opinion.”). To identify this problem is not to suggest that no agreement will be possible at the legislative stage on any issues. To the contrary, on some questions, agreement should be easy to reach. As earlier suggested, for example, we can expect legislators widely to agree that certain offenses (say, murder and rape) are serious offenses as we have understood this classification. The easy questions aside, however, many issues raised by the principles will generate reasonable disagreement even among parties deliberating in good faith. Is embezzlement of corporate funds a serious offense? What about stalking? Or selling drugs? Or using drugs? Even assuming agreement on which offenses qualify as serious, there is the further need for consensus as to the scale of the punishment that might be imposed consistent with the principles. Is a ten-year sentence for arson disproportionately harsh? What about a twenty-five-year sentence for a third-time armed robber? Under what circumstances, if any, would life in prison without parole be consistent with the parsimony principle? And on the basis of what standards are we even to decide these questions? Even parties who do not know the details of their own particulars and who are trying in good faith to determine answers to these questions consistent with the principles we have identified could reasonably offer different answers to these questions.

we cannot expect the empirical evidence and analysis to be unassailable. It is instead inevitable that at least some policy decisions will rest on inadequate or faulty empirical evidence, speculative judgments, failures of communication, or mutual misunderstanding. The combination of reasonable disagreement and inevitably questionable evidence and analysis means that even were our legislators to deliberate in good faith as to the policies the principles require, we could never be certain that the policies agreed to under these circumstances would in fact be consistent with the demands of the principles. The inevitable imprecision of such judgments once real world factors and evidence are introduced means that "[o]ften the best we can say of a law or policy is that it is at least not clearly unjust."²⁷² This inevitable uncertainty is what gives the imposition of every criminal punishment in any liberal democracy its tragic character.²⁷³

If we cannot be certain as to the legitimacy of punishment even after such an undertaking as the foregoing, on what basis ought we to think that a Rawlsian theory of punishment would represent an improvement over the current system? For in the current system, too, there is disagreement over the validity of the policies the legislative process yields, and here too we cannot be confident of the legitimacy of the outcomes of this process.

We have, however, already noted one such contribution: we now at least understand the necessary character of legitimate punishment. We have thus narrowed and focused the scope of possible disagreement on this extremely contentious question. Furthermore, in

272. TJ, *supra* note 14, at 199.

273. Honig would view such uncertainty regarding the ultimate legitimacy of the imposition of punishment in any particular case as a strength of the theory. She argues that "a theory of justice should strive not to justify punishment so well that it is moved beyond the reach of politicization but to insist that although justification is always a part of the practice of punishment, it is never seamless, never complete." Honig, *supra* note 10, at 121. It is for this reason, in Honig's view, that punishment is always "a tragic situation, in Bernard Williams' sense: It is never simply the right thing to do. It is not something that we ever get right." *Id.*

specifying that legislators at the third stage must deliberate from behind a veil that screens out any knowledge of their own attributes and personal particulars or those of their voting base, we have further narrowed the range of arguments that may be introduced in support of any particular policy. On the model I offer, in order to protect their own interests and the interests of their particular constituents, legislators are forced by the veil to consider—on terms by now familiar—the implication of various policies from the vantage point of uncertainty.²⁷⁴ In this way, we have ensured due consideration for the possible effects of policy proposals on society's least powerful members as well as its most privileged. Even if we can do no more than narrow the scope of the disagreement in these ways, we will still have made a considerable advance over the situation which currently prevails.

D. American Criminal Justice Policy and the Demands of Legitimacy

There is, however, a further important contribution that can be made in the policy arena now that we are equipped with an understanding of the terms of punishment to which the parties would agree under fair deliberative conditions. That is, even recognizing the inevitability of reasonable disagreement among people of good faith as to many of the difficult questions the problem of punishment raises in modern complex societies, it is still possible to identify a number of policies currently prevailing in the United States which appear, in light of

274. Thus the question is no longer: what content would you—legislator or citizen—prefer to give to the criminal justice policies at issue, knowing all you do about who you are and your own likelihood of being either a target or a beneficiary of any criminal justice policies here selected? It is instead: what approach to particular problems of criminal justice would you conclude to be most consistent with the requirements of the principles of punishment identified in part IV above, knowing only the general facts about the workings of your society and nothing about your own particular circumstances, or the way the sentencing policies ultimately selected will affect you or your constituents?

the principles, likely to yield punishments of questionable legitimacy. In what follows, I raise such questions in the context of specific criminal justice policies. In so doing, I enter territory that, if more concrete than the preceding arguments, is nonetheless necessarily more speculative. Rather than aiming to convince, I aim here simply to raise questions—and doubts.

The policies I consider below can be divided into two categories. The first category includes those policies that, although not inevitably producing sentences at odds with the demands of the principles, create by their design serious obstacles or disincentives to focused inquiry into the punishment demanded in particular cases. Because the principles require such an inquiry, there is a strong reason for skepticism that the sentences such policies yield would satisfy the requirements of the principles. The second category includes those policies that, by their very terms, either directly conflict with the principles or create a very high likelihood that their implementation would do so. Such policies would, from the perspective of the principles, raise legitimacy problems on their face.

1. Obstacles to Careful Deliberation

Read together, the principles of punishment we have identified impose a particular obligation on legislators charged with deriving legitimate criminal justice policies from the principles behind the modified veil. They have a duty, that is, to undertake a searching inquiry into the considerations put in issue by the principles: the relative seriousness of particular offenses, the burden imposed by particular punishments, the deterrent effect punishments may be reasonably certain to achieve, and the appropriate basis on which to ground these judgments to ensure that the results are as legitimate as possible given non-ideal circumstances. In many cases, the answers to the questions legislators face in performing this task would admit of reasonable disagreement. But this does not mean that all legislative conclusions are necessarily reasonable.

In evaluating existing policies, we should be particularly concerned with those policies that are incompatible with the kind of careful inquiry required by any good faith attempt to develop sentencing schemes consistent with the demand of the principles. More specifically, we should be wary of any policy that tends to preclude policymakers from considering any sentences imposed in light of the demands of the parsimony principle. This principle requires that punishments be no greater than necessary to achieve an appreciable deterrent effect against crimes of a severity equal to or greater than the punishment imposed. If legislatures are to honor this principle, they must maintain the flexibility to rethink sentences in particular cases, or indeed to rethink the whole legislative approach to punishing certain offenses when the circumstances demand such reevaluation.

Of particular concern in this regard are omnibus mandatory sentencing schemes which lump together a range of offenses and prescribe the same minimum sentence for each. Such schemes preclude the possibility for the focused consideration of the characteristics of each offense and the likely deterrent effect of the prescribed punishment which application of the principles demands. Consider, for example, California's three-strikes law, which imposes a mandatory minimum sentence of twenty-five years when a defendant with two prior listed felony convictions (that is, two "strikes") is convicted of any other felony.²⁷⁵ Or consider the federal statutes establishing mandatory minimum sentences ranging from five years to life in prison without parole (LWOP) for a wide variety of drug-related offenses.²⁷⁶ Under the California scheme, offenses triggering a third strike and thus a twenty-five

275. California's three-strikes law is codified in two virtually identical statutory provisions: Cal. Penal Code §§ 667(b)-(i), 1170.12(a)-(d) (West 2003). When a defendant convicted of any felony is found to have been convicted previously of two or more "serious" or "violent" felonies—"strikes"—the statute imposes a mandatory minimum sentence of twenty-five years. See §§ 667(e)(2), 1170.12(c)(2)(A)(i)-(iii).

276. See 21 U.S.C. § 841 (b)(1)(A)-(C); 21 U.S.C. § 960(b) (2004).

year sentence can thus range from the most violent (say, murder) to the least serious (say, stealing some video tapes from K-Mart,²⁷⁷ or stealing three golf clubs from a pro shop²⁷⁸). And under federal sentencing provisions, although drug offenders are punished severely “when death or serious bodily harm results” from their offense,²⁷⁹ these serious sentences are not reserved for such circumstances. Under federal law, for example, defendants convicted of possessing with intent to distribute one kilo of heroin “or a mixture containing heroin,” five kilos of cocaine, fifty grams of crack or one thousand kilos of marijuana face a mandatory sentence of LWOP if the defendant has two prior felony drug convictions²⁸⁰—and any prior felony drug convictions will do.²⁸¹ Such sentences may arguably satisfy the demands of the parsimony principle where they are applied to repeat violent offenders or to murderers when such punishment may be shown reasonably certain to deter future murders—a claim that has at least intuitive appeal.²⁸² But neither the California three-strikes scheme nor the federal drug laws is limited to such cases. As a

277. See *Lockyer v. Andrade*, 538 U.S. 63 (2003) (upholding the constitutionality of a mandatory sentence of fifty years-to-life imposed pursuant to the California three-strikes law for two third-strike counts of petty theft with a prior, imposed when the defendant stole a total of nine videotapes from K-Mart stores).

278. See *Ewing v. California*, 538 U.S. 11 (2003) (upholding the constitutionality of a mandatory sentence of twenty-five years-to-life for a third-strike grand theft conviction, imposed when the defendant stole three golf clubs from a pro shop).

279. For a defendant convicted under 21 U.S.C. § 841(b)(1)(A)-(C) (2003) of manufacturing, distributing, or dispensing a controlled substance, or possessing a controlled substance with intent to do any of these things, the base sentence is a life term when death or serious bodily injury results and the defendant had at least one prior conviction for a similar offense. See United States Sentencing Commission, Guidelines Manual, §§ 2D1.1(a)(1), 5A (Nov. 2002).

280. See 21 U.S.C. § 841(b)(1)(A). This is effectively a “three-strikes and you’re out” provision for drug offenders.

281. See *id.*

282. This is not to say that legislators may make such judgments based on mere intuition. I mean merely to suggest that such a sentence may well on further investigation prove to satisfy the demands of the parsimony principle and thus is not among those policies the legitimacy of which we are currently in a position to question.

result, offenders may well end up spending the rest of their lives in prison when their doing so would create no further protection for the higher-order interests of other members of society and is thus gratuitous.

Mandatory sentencing policies on this scale—or, indeed, on any scale—create a further obstacle to imposing sentences consistent with the parsimony principle: their mandatory character forecloses the possibility of reassessing whether ongoing incarceration would indeed serve the deterrent purposes this principle requires. Policies that operate in this way stand to violate the parsimony principle regardless of the animating deterrence theory. Certainly, the reasons why the parsimony principle might demand the shortening of a previously imposed sentence would vary with the deterrence theory on which the initial punishment was based. Where, for example, the calculation was based on specific deterrence or incapacitation, the parsimony principle would require the state to release an offender once it was determined that he or she posed no further threat to the security and integrity of innocent others. Under such circumstances, mandatory sentences that foreclose the possibility of revisiting the case to determine whether the anticipated threat is indeed still ongoing are thus certain to run afoul of this obligation. And mandatory sentencing schemes are a threat in this regard even where the initial punishment was fixed on a theory of general deterrence which made the ongoing threat posed by a particular offender irrelevant to the determination, for such policies still foreclose the reconsideration of sentences in light of emerging evidence. Where the state of the evidence changes over time to suggest that a particular sentencing policy overstated the length of sentence likely to yield a reasonably certain appreciable deterrent effect, the parsimony principle would require reconsideration of sentences crafted under that discredited or otherwise questionable theory. Furthermore, given that general deterrence claims are necessarily the more speculative, a good faith commitment to realizing the demands of the parsimony principle would necessarily

require sentencing policies that are flexible enough to take account of changing understandings. But whatever the underlying deterrence rationale, where mandatory minimum sentences are imposed by statute in a blanket way, there is a real danger that resulting sentences will violate the parsimony principle.

It is not only mandatory sentencing schemes that are potentially problematic in these ways, but also any policies that create incentives to maintain a prison population of a certain size, irrespective of the effect of doing so on the security and integrity of any citizen. And in this regard, of perhaps greatest concern are policies that create the potential for parties with political power or political connections to profit financially from large-scale incarceration. Under such conditions, the danger exists that the parties will make sentencing decisions, not in order to satisfy the concerns of the parsimony principle, but instead to expand the prison population and thus the potential for personal profit. Two current phenomena are of particular concern in this regard: the extensive political power of prison guard unions, whose members have a direct financial stake in maintaining and expanding the prison population;²⁸³ and the increasing trend toward delegating the management of public penal facilities or whole inmate populations to private for-profit corporations,²⁸⁴ whose continued profits and long-term

283. See Fox Butterfield, *Study Calls California Parole System a \$1 Billion Failure*, N.Y. Times, Nov. 14, 2003, at A24 (“[California’s prison] guards and parole officers have a financial incentive to keep the number of inmates high, helping preserve their jobs and ensure high salaries.”); Dan Morain & Jenifer Warren, *Battle Looms over Prison Spending in State Budget*, L.A. Times, Jan. 22, 2003, at 1 (noting that the “26,000-member prison guards union . . . is among the biggest campaign donors in California, giving \$3.4 million to [California Governor Gray] Davis directly and indirectly since his first run for governor in 1998, including more than \$1 million last year alone”); Matthew Heller, *Delano’s Grand Illusion*, L.A. Times, Sept. 1, 2002, pt. 9 (Magazine), at 8 (discussing Davis’s decision to budget \$335 million for a new prison in 1999, after the prison guards officers union contributed \$2 million to his campaign the year before).

284. At least twenty-seven state governments, as well as the federal government and some localities, contract out prison operations to private companies, which as of 1995 housed over 80,000 prisoners across the country, and today have the capacity for over 130,000 inmates. See Fox Butterfield, *For*

future also depend on a steady and continuous growth in the prison population.²⁸⁵ In jurisdictions where either of these conditions obtains, the danger exists that political influence²⁸⁶ will produce both attitudes and policies that favor longer sentences and disfavor parole, irrespective of whether these approaches and the sentences they yield can be shown to have any appreciable deterrent effect on the commission of serious offenses.

2. The Authorization of Illegitimate Practices

The second category of policies raising legitimacy questions includes those policies that by their very terms are either directly in conflict with the demands of the principles, or create a high likelihood of such violations. It is not only policies that specify the statutory penalty for particular offenses that raise legitimacy questions when viewed from the perspective of the principles. For, in addition to parsimony in sentencing, the parties would also

Privately Run Prisons, *New Evidence of Success*, N.Y. Times, August 19, 1995, at A7; Nzong Xiong, *Private Prisons: A Question of Savings*, N.Y. Times, July 13, 1995, at C5; Ted Strickland, *Private Prisons: The Bottom Line*, Wash. Post, June 13, 1999, at B1.

285. Contracts between state Departments of Corrections and private prison providers are structured on a cost-plus basis: the firm gets a set payment per inmate per day in exchange for assuming responsibility for running the facility and providing for inmates' needs. From the private provider's perspective, therefore, more inmates mean more money. There are, however, dangers for inmates in this arrangement, since "the opportunity for private profit is found only in the ability of the contractor to deliver the agreed services at a cost below the negotiated sum." Richard W. Harding, *Private Prisons and Public Accountability* 2 (1997).

286. See Steven Donziger, *The Prison-Industrial Complex: What's Really Driving the Rush to Lock 'Em Up*, Wash. Post, Mar. 17, 1996, at C3 (explaining that because private prisons are "funded entirely by government, firms like [Corrections Corporation of America, the nation's largest private prison management firm] need to ally themselves with politicians to sustain their growth"). In Tennessee, for example, an investigation into the Nashville-based CCA following announcement of a bill to privatize all 21 of the state's prison revealed "a small but impressive network of political contacts, a history of generous campaign contributions by CCA executives, and business ties among [CCA owner Tom] Beasley and top state officials." Richard Locker, *Personal, Political, Business Ties Bind CCA*, State, Com. Appeal, May 25, 1997, at B3.

require of the state two further guarantees: that meaningful protections be established against wrongful conviction, and that, at the very least, the state may not have recourse to punishments involving inhumane conditions of confinement where an alternative punishment would satisfy the demands of the parsimony principle. Judging the American criminal justice system in its current form on the basis of these standards, it seems clear that there is a considerable gap between existing policies and the policies we could expect of legislators deliberating under suitable standards of impartiality.

On the state's obligation to provide meaningful protections against wrongful conviction, consider the current status of state funding for indigent defense.²⁸⁷ In an adversary system such as that on which the American criminal justice system is based, the provision of effective counsel serves a number of goals. Central among these is protection against the wrongful conviction of a falsely accused defendant. In practice, however, indigent defense is severely underfunded across the country, with the effect that a great many defendants lack adequate representation, even when facing the most serious charges.²⁸⁸ And this is so despite growing levels of budgetary support for prosecutors.²⁸⁹ This underfunding

287. See, e.g., W. Andy Hardin, *What Price Is Justice?*, 37 *Tenn. B.J.* 23, 24 (2001) (explaining that "[s]ince its creation in 1989, the [Tennessee] District Public Defenders Conference has struggled with inadequate staffing and excessive caseloads" and that in its first ten years of operation, the public defenders' office grew by only 63%, while the caseload for public defenders grew by 276% in the same period).

288. The high cost of mounting an adequate defense was starkly illustrated during the O.J. Simpson trial, but evidence of the investment frequently required to defend oneself against state criminal charges is hardly in short supply. In one such case, which received international attention, Spanish citizen Joaquin Jose Martinez spent years on Florida's death row and only managed to demonstrate that "the prosecutor's evidence was full of holes" after his parents raised half a million dollars to cover the legal fees of private attorneys. See Suzanne Daley, *President Facing Skeptical Europe on Trip this Week*, *N.Y. Times*, June 11, 2001, at A1. The case prompted "an array of headlines" in the Spanish papers "like 'Paying not to die.'" *Id.*

289. See Richard S. Frase, *The Search for the Whole Truth about American and European Criminal Justice*, 3 *Buff. Crim. L. Rev.* 785, 810 (2000) (reviewing

has been facilitated by judicial interpretation of the Sixth Amendment right to counsel, which the Supreme Court has interpreted so narrowly²⁹⁰ that courts routinely deny inadequate assistance of counsel claims even where the evidence suggests not only serious incompetence but even incapacitation on the part of counsel.²⁹¹ Yet to the extent that the states are guided by this limited constitutional standard and under-invest in indigent defense, they create a serious risk of wrongful conviction that directly conflicts

William T. Pizzi, *Trials Without Truth: Why Our System of Criminal Trials Has Become an Expensive Failure and What We Need to Do to Rebuild It* (1999)) ("Most persons charged with crime are indigent, and public funding of defense staffing and investigative expenses (e.g., private investigators) consistently lags behind funding of police and prosecution services, which enjoy greater political popularity."); Kim Taylor-Thompson, *Effective Assistance: Reconceiving the Role of the Chief Public Defender*, 2 J. Inst. Stud. Legal Ethics 199, 221 n.12 (explaining that in 1990, of the \$74 billion spent on the justice system by federal, state, and local governments, "42.8% went to police, 33.6% to corrections, 12.5% to the courts, 7.4% to prosecution . . . [and] 2.3%" to "[p]ublic defense") (citing Richard Klein & Robert Spangenberg, *The Indigent Defense Crisis* 1 (1993) (prepared for the American Bar Association Section of Criminal Justice Ad Hoc Committee on Indigent Defense Crisis)); see also *id.* at 206 n.26 ("[T]he nation's spending on defense services pales in comparison to the expenditure on prosecution and enforcement.").

290. See *Strickland v. Washington*, 466 U.S. 668, 690 (1984) ("A convicted defendant making a claim of ineffective assistance must identify the acts or omissions of counsel that are alleged not to have been the result of reasonable professional judgment. . . . [T]he court should recognize that counsel is strongly presumed to have rendered adequate assistance and made all significant decisions in the exercise of reasonable professional judgment.").

291. See, e.g., *United States v. Muyet*, 994 F. Supp. 550, 560 (S.D.N.Y. 1998) (sleeping attorney); *United States v. Rondon*, 204 F.3d 376, 381 (2d Cir. 2000) (attorney disbarred by state bar during defendant's trial in federal court); *Hernandez v. Wainwright*, 634 F. Supp. 241, 245 (S.D. Fla. 1986) (alcohol consumption during trial); *Berry v. King*, 765 F.2d 451, 454 (5th Cir. 1985) (drug use during trial); *Dows v. Wood*, 211 F.3d 480, 485-86 (9th Cir. 2000) (defense attorney suffering from Alzheimer's); *Smith v. Ylst*, 826 F.2d 872, 876 (9th Cir. 1987) (defense attorney suffering mental illness). See also Stephen B. Bright, *Counsel for the Poor: The Death Sentence Not for the Worst Crime but for the Worst Lawyer*, 103 Yale L.J. 1835, 1842-43 & nn.49-55 (1994) (listing cases in which defense attorneys were ignorant of relevant law, "presented conflicting defense for the same client, referred to their clients by a racial slur, cross-examined a witness whose direct testimony counsel missed because he was parking his car, slept through part of the trial, . . . [were] intoxicated during trial," or filed incomplete or minimal briefs (citations omitted)).

with the terms of legitimate punishment as we have construed them here.²⁹²

Or consider, in light of the state's obligation to avoid gratuitous inhumane punishments, the conditions of confinement facing inmates at prisons and jails across the country, which strongly suggest that this requirement is routinely being violated.²⁹³ The widespread incidence of rape and sexual assault in prisons and jails and the ongoing threat of such abuse, which is a permanent aspect of incarceration at many prisons, would alone serve to prove the point.²⁹⁴ However one would understand the category of inhumane punishments, the conception would have to reach conditions in which inmates live with the ongoing threat of rape or other forms of sexual abuse, a threat that is frequently realized, if not against oneself then against some substantial number of one's fellow inmates. Yet it is not an exaggeration to say that this describes the conditions of confinement of many of the

292. Perhaps the most extreme illustration of this danger is the current practice in some southern counties of contracting out indigent defense to the lowest bidder. See Amy Bach, *Justice on the Cheap: For Many Indigent Defendants, the Right To a Lawyer Doesn't Mean Much*, *The Nation*, May 21, 2001, at 25 (highlighting the growing use, especially in Southern states, of low-budget contract systems to replace court-appointed systems); *Miscarriages of Justice Abound for Georgia's Poor*, *Atlanta J. & Const.*, Apr. 23, 2001, at 8A (describing one particular contract attorney who entered guilty pleas "negotiated in the hallways" for nearly all of 94 defendants represented).

293. Granted, to know for certain whether this is so, we would need to analyze each sentence in light of the evidence regarding the prospects for deterrence. But given the broad range of offenses for which individuals are incarcerated, and in particular the extent of incarceration for non-violent offenses, it seems reasonable to assume that the inhumane conditions of confinement that currently prevail at many penal facilities would not be justifiable in terms of the principles.

294. It might be argued that such conditions of confinement do not amount to an explicit policy of the sort here discussed, but only a regrettable, unavoidable side-effect of other problems of the criminal justice system. But if our concern is with the nature of state punishment and the burden it imposes on individual convicted offenders, there are no grounds for treating the practical effects of affirmative actions differently than the practical effects of neglect and indifference to the conditions of confinement under which sentences are carried out. The conditions under which a convicted offender is confined are as much a part of the punishment as the duration of the official sentence. They are therefore as much a concern of the principles of punishment as official sentencing policies.

nation's inmates.²⁹⁵ And even assuming cases in which inhumane conditions of incarceration were found necessary to deter the commission of inhumane crimes, it is unlikely that it would be found necessary on this basis that such conditions represent a perpetual part of the offender's sentence. In any case, with over 50% of incarcerated offenders having been convicted of non-violent offenses,²⁹⁶ it is unlikely (if not impossible) that an ongoing threat of rape or sexual assault would be found consistent with the demands of the principles when applied to most or all of these cases.

To take just one other example of conditions that would raise serious concerns in this regard, consider the current state of prison overcrowding. Expert estimates of the space "a long-term inmate must have to himself . . . to avoid serious mental, emotional, and physical deterioration" range from fifty to eighty square feet²⁹⁷—an area, as Justice Marshall once pointed out, that is "smaller

295. See Eli Lehrer, *Hell Behind Bars*, Nat'l Rev., Feb. 5, 2001, at 24 ("Prison rape may be American's greatest ignored crime problem."). According to Human Rights Watch, even the most conservative estimates indicate that "more than one in ten inmates in prisons surveyed was subject to sexual abuse." Human Rights Watch, *No Escape: Male Rape in U.S. Prisons* 5 (2001). Yet the chronic under-reporting of rape by prison inmates suggests a much higher rate even than this. See *id.* at 132 (citing one study that "found that only 29 percent of victimized inmates had informed prison officials of the abuses they suffered," and another that "found that of an estimated 2,000 rapes that occurred [in Philadelphia facilities], only ninety-six had been reported to prison authorities"). Sexual abuse among incarcerated populations is also difficult to quantify because it does not always appear to take the form of forced intercourse. See *id.* at 6-7; Human Rights Watch, *All Too Familiar: Sexual Abuse of Women in U.S. Prisons* 1-2 (1996); James Gilligan, *Violence: Our Deadly Epidemic and Its Causes* 165 (1996). This observation holds true for female inmates as well as male inmates, as male corrections officers may "use[] their near total authority [over female inmates] to provide or deny goods and privileges . . . to compel them to have sex . . ." Human Rights Watch, *All Too Familiar*, *supra*, at 1.

296. See Vincent Schiraldi et al., *Justice Policy Inst., America's One Million Nonviolent Prisoners* (1999), http://www.cjcj.org/pubs/one_million/one_million.html. To put this in some perspective, the number of nonviolent offenders incarcerated in the United States is larger than the combined populations of Wyoming and Alaska. Jason Ziedenberg & Vincent Schiraldi, *Justice Policy Inst., The Punishing Decade: Prison and Jail Estimates at the Millennium* 3 (1999).

297. See *Rhodes v. Chapman*, 452 U.S. 337, 371 (1981) (Marshall, J., dissenting).

than that occupied by a good-sized automobile.”²⁹⁸ Yet overcrowding in the nation’s penal facilities,²⁹⁹ an artifact of the dramatic growth in the prison population over the last three decades,³⁰⁰ has meant that in practice, inmates in many prisons and jails across the country are double-celled in cells no bigger than the area described by Justice Marshall. Not only does double-celling in such close quarters exacerbate the threat of rape as well as other forms of violence and coercion,³⁰¹ but such conditions also give rise to the forms of deterioration just noted, depriving inmates of the minimum physical space humans need to preserve a sense of self. To this extent, prison conditions as they currently exist across the country create the real possibility of undermining not only the security and integrity of inmates but also the very bases of their moral personhood. And if this is the case, it is very possible that many if not most prison inmates currently behind bars face

298. *Id.* Federal standards call for single occupancy for “rooms, cells and cubicles” between fifty-five and seventy-five square feet, depending on security designation. Federal Bureau of Prisons, U.S. Dep’t of Justice, Program Statement 1060.11, *Rated Capacities for Bureau Facilities* (1997), http://www.bop.gov/progstat/1060_011.pdf. High security facilities are not supposed to double-cell in an area less than seventy-five square feet, although in minimum security facilities all cells of fifty-five square feet or more are to be used for double-celling. See *id.* In California, regulations mandate a minimum of sixty to seventy square feet of “floor area,” depending on the facility’s security classification. See 24 California Board of Corrections Regs. § 470A.2 (2001), http://www.bdcrr.ca.gov/regulations/t-24_adult_regs_final/t24-470a2.htm#470A.2.6. California permits double-celling in cells of sixty square feet even in the highest security facilities. See *id.*

299. As of January 1996, thirty-six states and the District of Columbia were under court order to reduce overcrowding in some or all of their prisons. See National Prison Project, *Status Report: State Prisons and the Courts* 1 (1996). By the end of 2000, 310 prisons nationwide were operating under such a court order, and “courts had placed population caps on forty-four prisons.” Lynn S. Branham, *Cases and Materials on the Law of Sentencing, Corrections, and Prisoners’ Rights* 621 (6th ed. 2002). And by the end of 2001, thirty-three states, the District of Columbia, and the federal system were housing prisoners in jails and other facilities because of overcrowding. See Harrison & Beck, *supra* note 4, at 8.

300. See *supra* pp. 310-11 and note 5.

301. See Wilbert Rideau, *The Sexual Jungle*, in *Life Sentences* 73 (Wilbert Rideau & Ron Wikberg eds., 1980) (describing a system among inmates in which dominant prisoners “owned” the bodies of the weaker prisoners, and “loaned” them out for sex to other inmates in exchange for money, drugs, cigarettes, or other goods).

conditions of confinement that qualify as inhumane, in direct conflict with the requirement of the principles.

One obvious rejoinder to this critique is that the state officials responsible for the policies just described make no claim to be abiding by the principles identified here. It is thus no surprise if there is a tension between the content of these policies and the demands of the principles produced by our analysis. Yet the fact that the policies at issue sprang from a different normative basis does not provide an adequate defense of their legitimacy. If I have adequately made the case that the principles the parties would select under the deliberative conditions modeled by the Rawlsian framework represent the terms of legitimate punishment in liberal democracy—or even simply that they represent terms of punishment of greater legitimacy than those produced by our current legislative process—then the principles represent an appropriate basis from which to judge the policies currently in force.

And whatever the limitations of the present analysis, I am confident that the deliberative conditions of the original position—and even those of the legislative stage, with its modified veil of ignorance—produce principles of punishment of greater legitimacy than those embodied in many of our current criminal justice policies. To see that this is so, one need only employ a simple thought experiment predicated on the animating principle of the veil of ignorance. That is, ask yourself: if you were charged with developing criminal justice policies for your society as if you really didn't know which social position you would end up occupying once the veil was lifted, would you really opt for the policies and conditions just canvassed?

VI. CONCLUSION

In this article, I have applied Rawls's deliberative framework to the problem of legitimate punishment in liberal democracy. Rawls did not intend his model to apply to problems of partial compliance. I have, however, sought to show that by modifying certain assumptions regarding

Rawls's well-ordered society, the ideas of the original position and the veil of ignorance can be made to yield insights directly relevant to the problems of our non-ideal world.

My use of this model is motivated by a simple idea: to be legitimate, the exercise of the state's power to punish criminal offenders in a liberal democracy must be consistent with principles the terms of which all members of society would accept even if they did not know where in the criminal justice hierarchy they would turn out to be. This requirement, which amounts to a strict standard of impartiality, is in one sense the enactment of the basic liberal ideals of moral equality and individual sovereignty. By screening out the influence of morally arbitrary considerations on deliberations as to the appropriate scope of state power, the model accords the perspectives of all members of society equal consideration and respect, and affirms the status of each person as simultaneously a fellow human being with his or her own conception of the good and a fellow citizen to whom the exercise of state power must be justified. In this way, the framework we have employed is a quintessentially liberal one, as is the theory of punishment we have derived from its application.

Equal consideration and respect, however, do not necessarily require equal treatment. As the foregoing analysis has shown, the liberal democratic state may, consistent with its own normative commitments, single out certain members of society and subject them to the punishment of incarceration. Certainly, its power to do so is not without limits, and in this article, I identify five principles with which state punishment must be consistent if it is to be legitimate.³⁰² But within the limits these principles represent, the arguments contained herein have shown that the state may incarcerate convicted offenders—

302. At the core of the principles I identify is the parsimony principle, the basic idea of which is that the punishment of convicted offenders must be no more severe than necessary to yield an appreciable deterrent effect on the commission of serious offenses. For the complete list of these principles, see above, *supra* pp. 408-09, 411.

even for extended periods—without exceeding its legitimate authority.

Identifying the principles with respect to which state punishment must be consistent to be legitimate is not, of course, to guarantee the legitimacy of all punishments actually imposed on convicted offenders. Even assuming the good-faith efforts of policymakers to design criminal justice policies consistent with the principles, the inevitability of reasonable disagreement over what punishments the principles allow under a given set of circumstances means that, in practice, we can never be fully confident of the legitimacy of any punishments imposed. Still, as I argue in part V, the principles of legitimate punishment identified herein do provide the basis for calling into question the legitimacy of a range of criminal justice policies currently in force in the United States, including mandatory minimums, California's three-strikes law, the underfunding of indigent defense, and the widespread sexual violence and overcrowding in the nation's prisons and jails. In this way, notwithstanding the many and obvious obstacles to full implementation of the principles here identified, the theoretical analysis I offer nonetheless provides a basis for challenging the legitimacy of a great many criminal sentences being served by fellow citizens and fellow human beings in American prisons right now.