TOWARD A POLITICAL THEORY OF CRIMINAL LAW: A CRITICAL RAWLSIAN ACCOUNT

*Emmanuel Melissaris**

The aim of this paper is to outline a political theory of criminal law, that is, a theory that does not rely on any controversial moral view on fault and punishment. The argument is based largely on John Rawls's work but also addresses some inconsistencies regarding crime and punishment therein. The paper will make a case for a mixed theory of criminal law and punishment along the following lines: The fair terms of social cooperation generate duties for the violation of which one can be held properly responsible. However, nothing in this determines what the accountability-seeking measure may be or its intensity. These are matters of appropriateness to be determined in terms of the assurance of the participants in a well-ordered society and the long-term stability of social cooperation. Criminalization and punishment are contingent, historically qualified means of achieving stability and assurance, serving only as last resort. Although the institution of punishment is subject to the constraints of the rule of law, which stem directly from the liberty principle, questions such as the intensity of punishment or the proportionality between offenses and penalties depend on the proper workings of the utilitarian calculus in combination with the requirements of democratic decision-making.

Keywords: *criminal law, punishment as political, John Rawls, mixed theories of punishment*

INTRODUCTION

This paper builds on two fundamental liberal assumptions: first, political and legal institutions must be respectful of everyone's status as free and equal;

*London School of Economics, Law Department. Many thanks to Nicola Lacey, Neil Duxbury, Mike Redmayne, and Robert Reiner for comments on previous drafts. I am also indebted to the journal's reviewers. I am entirely responsible for any remaining mistakes.

New Criminal Law Review, Vol. 15, Number 1, pps 122–155. ISSN 1933-4192, electronic ISSN 1933-4206. DOI: 10.1525/nclr.2012.15.1.122.

secondly—and this is the procedural counterpart to the substantive requirement of liberalism—political institutions can only be justified in political terms. This means that political justification cannot invoke controversial moral doctrines, which inescapably generate irresolvable disputes. Political institutions must be grounded in a manner that is neutral and respectful of conflicting moral convictions while recognizing that it is necessary for such belief systems to be coordinated and to coexist peacefully.

The institutions of criminal law and punishment are no exception to these fundamental liberal requirements. This paper's aim is to make a contribution to the discussion of how the criminal law can be accommodated in a substantive and procedural liberal political theory. More specifically, I will explore the place of the criminal law in one of the most powerful and influential versions of substantive and procedural liberalism, namely John Rawls's theory of the state, law, and justice.

The discussion of criminal law and punishment in Rawls as well as in the greatest part of the literature on distributive justice is conspicuously underdeveloped. As a result, the precise relationship between the liberal state, distributive justice, and punishment remains rather vague and urgently calls for clarification, not least because an account of justice is incomplete without an account of criminal responsibility. This is not to suggest that criminalization and punishment are a priori ingredients of modern political communities; in fact I argue quite the opposite in this paper. But because wrongdoing is widely viewed as necessitating punishment and because criminal justice is so central an institution in contemporary political communities, indeed perhaps too much so, it is important to put it in the right normative perspective to debunk myths and make it compatible with political justice.

Constructing a Rawlsian theory of criminal law will at times entail clarifying and qualifying some of Rawls's arguments on punishment. If on occasion I take a critical stance and depart from Rawls's text, I hope that my arguments are true to the point of his thought. I also hope that my argument here will go some way to completing the Rawlsian project of constructing a system of fair social cooperation.

In the first part I will briefly summarize the contours of the Rawlsian theory of distributive justice and Rawls's passing comments on criminal law and punishment. Rawls seemed to view punishment in largely consequentialist terms as aiming at maintaining stability and offering assurance. However, a fully fledged utilitarian view is clearly out of place in the Rawlsian scheme, both because it entails subscribing to a comprehensive moral doctrine and

because it is not sufficiently respectful of personhood. At the same time, there are some seemingly strong retributivist undertones in Rawls's view of the criminal law. First, he seems to narrow down the latter's scope to a specific category of wrongs, which relate to what he terms "natural" duties. This can be, and has been, read as implying that one is *criminally* responsible for and that punishment is the *right* response to the violation of these specific duties. Secondly, Rawls seems to draw a close link between desert and punishment. Once again, neither position coheres with the rest of the Rawlsian scheme of justification of the state and the principles of justice. First, retributivism of all kinds inescapably relies on some controversial doctrine therefore being excluded from public reason. Secondly, having already rejected preinstitutional desert as a ground of justice, it appears incongruous to rely on precisely that type of desert for the justification of the criminal law.

I will therefore try to construct a Rawlsian theory of the criminal law by providing answers to two questions: how responsibility can be reconciled with the largely consequentialist need for stability and assurance; and how the rejection of desert as a ground for allocation of social advantages can be reconciled with Rawls's acceptance of desert as a ground for retribution. In the course of addressing these problems, I will argue that in the Rawlsian scheme, the criminal law and punishment can only be regarded as contingent institutions providing appropriate, as opposed to a priori morally or otherwise right, responses to the violations of political duties on the basis of the offender's responsibility. The "appropriateness" of criminalization and punishment is to be judged in view of such facts as the deterrent effect of punishment, the state of citizens' motivational disposition, and so forth. Such a "mixed" theory successfully overcomes the problems of both retributivism and fully fledged utilitarianism and steers clear of any metaphysically or morally laden correlations between wrongs and punishment.

I. RAWLS'S THEORY OF DISTRIBUTIVE JUSTICE AND THE OBITER REFERENCES TO PUNISHMENT

It is clearly not possible to give a full account of the Rawlsian theory of distributive justice in this context. It is, however, necessary to outline some of its foundational premises and relate them to the Rawlsian theory of criminalization and punishment.

Rawls's aim is impartially to construct a theory of justice, which will be serviceable in the context of modern, constitutional, democratic societies characterized by the "fact of reasonable pluralism."[1] The crucial question is how it may be possible to establish fair terms of social cooperation, which will be neutral between and respectful of conceptions of the good and will therefore be acceptable by all irrespective of their belief systems and the comprehensive moral doctrines to which each may subscribe.

The first step in this process of construction is the counterfactual *original position* (hereinafter OP). People who find themselves in the OP are behind the "veil of ignorance," which leaves out of the process any conceptions of the good that they may hold and any personal characteristics, beliefs, and preferences. Apart from some general facts about human societies, such as the scarcity of resources, the only available information to them is: (a) that they are rational and reasonable, the former referring to their ability to correlate means and ends, and the latter to their ability to propose or accept principles conducive to fair social cooperation; and (b) that they regard themselves and each other as *free* and *equal*. Their equality consists in that they all regard themselves and each other "as having to the essential minimum degree the moral powers necessary to engage in social cooperation over a complete life and to take part in society as equal citizens."[2] They are considered free in that they are capable of holding conceptions of the good and in their capacity as "self-authenticating sources of valid claims."[3]

On the basis of these traits and moral powers, people in the OP will *agree* on principles of social cooperation, which are acceptable by all, because they will satisfy the participants' desire to pursue their self-interest without losing sight of their regard for others, the cooperation with whom is a precondition for the satisfaction of one's own interests. The two principles of justice that Rawls envisions people in the OP to agree on are, famously: (1) the liberty principle—all should enjoy the most extensive (or, according to a later formulation, a fully adequate[4]) scheme of rights and basic liberties; (2a) all should enjoy fair equality of opportunity; and (2b) the difference principle—all should enjoy fair equality of opportunity; inequalities are permissible only

1. I generally refer to the final formulation of Rawls's theory of justice, i.e., after Political Liberalism (2005) (1993); Justice as Fairness: A Restatement (E. Kelly ed., 2001) (hereinafter A Restatement).

2. Rawls, A Restatement, supra note 1, at 20.

3. Ibid. at 22.

4. Rawls, Political Liberalism, supra note 1, at 331.

to the extent that they benefit the worse-off members of society. The difference principle provides an incentive to the better-off to be more productive and earn more, while at the same time placing market-independent limits to profit.[5]

It cannot be overemphasized that these principles of justice are *political*. They are not grounded in or impose any comprehensive moral doctrine, whether religious or secular. The claim is that they can establish fair terms of social cooperation, which are neutral between such conceptions of the good and moral doctrines. One of the upshots of their political character is that they only apply to the political and not the private or the social spheres. Rawls has a narrow understanding of the political as comprising only the *basic structure* of society, that is, the set of political and economic institutions setting the framework of social cooperation, establishing the "background justice," to use Rawls's terminology.[6] Not wanting to overprescribe the content of the basic structure and in order to allow for flexibility in light of particular circumstances in various contexts, Rawls only offers us a general outline of what the basic structure may include: constitutional essentials; procedural rules regulating legislation and access to justice; specific rules establishing duties on the part of individuals.[7]

Now, if it were certain that all would respect and abide by the principles of justice without remainder, it would not be necessary to enforce them institutionally. If there were no distance between normative and behavioral expectations, then there would be no need to employ the coercive powers of the state. But this is clearly not the case in modern, complex societies. So, in the face of societal complexity and the uncertainty caused by the fact that not all are guaranteed to act on the right reasons stemming from justice, the law is meant to offer people assurance that the terms of reciprocity will be abided by and that all will have an external incentive to respect the principles of justice with everything that they require on a micro level.

---

5. J. Rawls, A Theory of Justice (1971).

6. Many of Rawls's critics have taken issue this. For instance, G.A. Cohen argued that a society, in which the principles of justice are not applicable to all in their everyday economic activity, and which lies outside the scope of the basic structure, cannot be just. G.A. Cohen, Rescuing Justice and Equality (2008). See also J. Habermas, Reconciliation Through the Public Use of Reason: Remarks on John Rawls's Political Liberalism, 92(3) J. Phil. 109 (1995). And for Rawls's response: Reply to Habermas, 92(3) J. Phil. 132 (1995) (reprinted in the 2005 edition of Political Liberalism, supra note 1).

7. Rawls, A Theory of Justice, supra note 5, at 195–201.

This is not to say that Rawls envisions the basic structure as playing the part of the Hobbesian Leviathan state.[8] Stability is first meant to be guaranteed by the requirement of *overlapping consensus*, that is, the requirement that norms binding for all will cut across all comprehensive moral doctrines so they will be acceptable by all. Secondly, the assurance provided by the basic structure and the coercive intervention of the state is to be accompanied by, and indeed anchored to, a *sense of justice*. It is this sense of justice that can guarantee the long-term stability of a well-ordered, fair society. Rawls backs this idea with a complex and partly moral psychological argument, which I will not delve into here.[9] The gist of it is that, if the basic structure of society is just and people are appropriately brought up in it, they will develop the inclination to act on reasons of justice not out of fear of sanctions or purely on the grounds of self-interest (as in the Hobbesian worldview) but rather because this is how they will fulfill their moral powers of freedom and equality *both* rationally *and* reasonably.

Rawls, and indeed most theorists of distributive justice, have rather little to say on punishment and the criminal law. The emphasis is placed mostly on allocative principles rather than on whether and, if yes, why and how it is permissible for the state to intervene coercively in response to the violation of certain rights or rules.

For Rawls punishment is an institution primarily serving as a guarantee of assurance and stability. It is worth citing at length the relevant passages:

> It is reasonable to assume that even 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 role of an authorized public interpretation or rules supported by collective sanctions is precisely to overcome this instability. By enforcing a public system of penalties government removes

8. T. Hobbes, Leviathan or, The Matter, Forme and Power of a Commonwealth, Ecclesiasticall and Civil (Richard Tuck ed. 1996) (1651).

9. Rawls, Political Liberalism, supra note 1, at 140ff. For a critique of Rawls's view on stability and congruence between good and justice, see Brian Barry, John Rawls and the Search for Stability, 105(4) Ethics 874 (1995). For an interesting discussion of the parallel between subjecting one's preferences to justice, on the one hand, and love, on the other, which Rawls draws in A Theory of Justice, see S. Mendus, The Importance of Love in Rawls's Theory of Justice 29 Brit. J. Pol. Sci. 57 (1999).

> the grounds for thinking others are not complying with the rules. For this reason alone, a coercive sovereignty is presumably always necessary.[10]
>
> . . .
>
> [In] agreeing to penalties that stabilize a scheme of cooperation the parties accept the same kind of constraint on self-interest that they acknowledge in choosing the principles of justice in the first place. Having agreed to these principles . . . , it is rational to authorize the measure needed to maintain just institutions, assuming that the constraints of equal liberty and the rule of law are duly recognized.[11]

But assurance and stability do not suffice. Rawls seems to ground punishment on *some* sense of responsibility:

> . . . the purpose of the criminal law is to uphold *basic natural duties*, those which forbid us to injure other persons in life and limb, or to deprive them of their liberty and property, and punishments are to serve this end. They are not simply a scheme of taxes and burdens designed to put a price on certain forms of conduct and in this way to guide men's conduct for mutual advantage. It would be far better if the acts proscribed by penal statutes were never done. Thus a propensity to commit such acts is a mark of bad character, and in a just society legal punishments will only fall upon those who display these faults.[12]

Elsewhere in the *Theory of Justice* in the course of accounting for the rule of law, he correlates punishment with the liberty principle:

> I have maintained that the principles justifying . . . sanctions can be derived from the principle of liberty. . . . We also see that the principle of responsibility is not founded on the idea that punishment is primarily retributive or denunciatory. Instead it is acknowledged for the sake of liberty itself. Unless citizens are able to know what the law is and are given a fair opportunity to take its directives into account, penal sanctions should not apply to them. The principle is simply the consequence of regarding a legal system as an order of public rules addressed to rational persons in order to regulate their cooperation, and of giving the appropriate weight to liberty. I believe that this view of responsibility enables us to explain most of the excuses and defenses recognized by the criminal law under the heading of *mens rea* and that it can serve as a guide to legal reform.[13]

---

10. Rawls, A Theory of Justice, supra note 5, at 240.
11. Id. at 576.
12. Id. at 314–15.
13. Id. at 241.

So, this is the tentative picture of criminal law and punishment emerging from Rawls's writings: The fundamental thesis is that the justification of criminalization and punishment, like the justification of any other political institution, cannot rest on moral grounds, it cannot reflect a comprehensive moral doctrine. This has two implications. First, in the context of the public, political forum in modern, liberal, constitutional states, nothing can be taken to be criminal in se. In other words, acts are not to be included in the criminal law by virtue of some moral property of theirs that transforms them into a specific type of wrong. Of course, what, if anything, does end up falling within the ambit of the criminal law may be considered by some to be morally objectionable too. The crucial reason for criminalization, however, will not be this overlap with some citizens' moral beliefs but the fact that the political, social-contractual conditions are met.

A significant upshot of this is that it cancels out the distinction between criminalization and punishment. If one is committed to the idea that there is an essential difference between criminal and other types of wrongs, then the inclusion in the criminal law may be dissociated from the question of punishment. An act would then count as "criminal" on moral grounds, but whether it ought to be punished is an analytically separate question. If, however, such robust moral argument is discounted from the outset, criminalization and punishment become parts of the same scheme. The inclusion of an act in the criminal law does not follow from its moral demerit but is rather a response of a specific kind to specific acts, though still not a response motivated by controversial moral reasons. Punishment in the narrow sense—that is, depriving offenders of liberties or facilities, which they would normally be able to enjoy—may be a way of organizing this response, but it is certainly not the only way. It is, for instance, conceivable that labeling one as a criminal would be considered a sufficient response to one's act. In light of this, for the remainder of the article I will not draw any substantive distinction between criminalization and punishment.

Secondly, it follows from the claim to neutrality of the Rawlsian scheme that the reasons for punishing cannot be part of a comprehensive moral or metaphysical doctrine. This already rules out a strong retributivism, which locates the reason for punishing in the wrongful act itself. But it also entails the rejection of a full-blown utilitarianism.

As for the question of what does justify punishment, according to Rawls injunctive rules are permissible (at least) to the extent that they are necessary to

enforce what he calls "natural duties." The violation of these duties establishes the conditions of responsibility. In the second passage above, we are told that only those who violate natural rights may be punished. In the last passage, it appears that this constraint stems directly from the principle of liberty. In other words, those who violate certain norms *deserve* to be punished, as their act is an expression of their "bad character." The practice of punishment specifically is in turn justified by the need for assurance and stability.[14] The amount of punishment is to be regulated by the rule of law and the principle of liberty, which sets the requirement of proportionality or fittingness of the penalty to the crime.

Straightforward as all this may seem at first glance, there are two central issues that need to be addressed.

First, what is the precise relationship between the need for stability on the one hand and individual responsibility on the other? How can the two be reconciled given that they potentially pull in opposite directions? If punishment is to be justified solely on the grounds of stability, then we are a short step away from a fully fledged utilitarian theory of punishment, which is entirely out of place in the Rawlsian scheme not least because it is disrespectful of agents' personhood and it is a comprehensive moral doctrine. But such a fully utilitarian tack also sits uncomfortably with Rawls's dicta on the relationship between punishment and the liberty principle, which implies that there is some relation of proportionality between the amount of liberties of which one can be deprived in response to the wrong that one has done. In other words, it is suggested that there is some relation (a relation of desert) between responsibility and punishment, which seems to clash, at least at first sight, with assurance or deterrence as the overall aim of punishment. If, on the other hand, punishment were to be justified on the basis of individual responsibility, then Rawls would have to subscribe to a substantive and necessary connection between wrongdoing and punishment. Rawls comes close to doing so in drawing, however cursorily, a connection between criminalization and the violation of natural duties. But all this dangerously verges on admitting a substantive, comprehensive moral theory as complementary to the theory of justice. And this too would be incoherent.

---

14. This, as I shall show later on in this paper, squares with the distinction between the justification of a practice, on the one hand, and the justification of specific instances, on the other, which Rawls drew in his Two Concepts of Rules, 64 Phil. Rev. 3 (1955).

Secondly, and this problem is related to the previous one, Rawls's accounts of distribution and retribution are arguably inconsistent.[15] The principles of distributive justice are explicitly based on a rejection of any notion of desert. For Rawls, naturally, hence accidentally, acquired talents have no bearing on the allocation of resources. No prejusticial and preinstitutional notion of desert is salient to the question of fair distribution.[16] This does not amount to saying that everyone should not be given the chance to develop their talents. On the contrary, the principles of justice (and arguably especially the principle of fair equality of opportunity) will ensure that everyone has the chance so to do. When it comes to retributive justice though, Rawls is prepared to accept that punishment is to be allocated according to desert and in response to one's "bad character." There is indeed an asymmetry here, which threatens to drag the whole theory into incoherence. If people deserve to be punished for an act or, worse even, for some character trait of theirs, then denying that desert has moral weight when it comes to the distribution of goods becomes significantly more difficult, if not altogether impossible.

Now, none of this would be an issue, if no account of criminal law were required, and if the scope of the Rawlsian theory of justice could be complete even if it were narrowed down to the distribution of resources alone. But this cannot be so. As Rawls admits, a theory of justice must be adjustable to nonideal conditions of partial compliance. This is not to say that state punishment is rendered morally or otherwise necessary or a priori. The fact, however, that it is a practice so central in modern states and that it is a prima facie way of dealing with partial compliance means that it must be tackled and put in the right perspective. And this must be done coherently in a way that does not undermine the foundations of the whole edifice. Therefore, the challenge for Rawls is to find a way of accounting for punishment that will: (a) be respectful of and somehow follow from the moral powers and traits of citizens, which are foundational to constitutional democracies, without regressing into a comprehensive moral doctrine; and (b) reconcile the rejection of desert as a principle of distribution with its admission as a principle of punishment.

---

15. This is the problem that has been discussed the most in the literature. See S. Scheffler, Justice and Desert in Liberal Theory, 88 Cal. L. Rev. 965 (2000); E. Mills, Scheffler on Rawls, Justice, and Desert, 23 Law & Phil. 261 (2004); J. Moriarity, Against the Asymmetry of Desert, 37 Noûs 518 (2003).

16. Cf. D. Miller, Principles of Social Justice (1999).

In what follows I will suggest some possible solutions to these two fundamental problems and, in the course of doing so, construct a coherent view of punishment based on the Rawlsian scheme of justice.

II. RECONCILING RESPONSIBILITY WITH ASSURANCE AND STABILITY

This is the first problem I have identified so far: Rawls views the criminal law as a matter of assurance and stability in the face of the inability or refusal of some to respect natural rights stemming from the social contract agreed on in the OP. At the same time, however, Rawls insists that punishment can only be meted out on the basis of responsibility and in accordance with the liberty principle and the rule of law, which it informs. But how can these be reconciled? What is the relation between stability, a by and large forward-looking aim, and the conditions of individual accountability?

Let us take things from the beginning. According to Rawls, the agreement struck in the OP gives rise to some *natural* duties. Some such duties are justice-independent: the duty to mutual aid, the duty not to harm each other, and the duty not to inflict unnecessary suffering.[17] They are "natural" not in the sense that they are traceable to some universal moral order, but in that they are unconditional and independent of any voluntary acts or institutional structures. As such they bind all and are owed to all, although they remain *political*.

At the same time, they do not flow exclusively from the principles of justice. This would amount to arguing that, say, violations of others' bodily, sexual, and psychological integrity are violations of the liberty principle, that the wrongfulness of such acts as murdering, raping, or assaulting consists in depriving others of their fair share of the corresponding liberties or primary goods. This, however, obviously misreads and depreciates what is at stake in such cases.[18] The acts of murder, rape, and assault turn against the victim as a political agent free, equal, and capable of social cooperation. The duty to abstain from such acts therefore emanates directly from the OP rather than

17. Rawls, A Theory of Justice, supra note 5, at 114.

18. For this argument see A. Duff, Penal Communications: Recent Work in the Philosophy of Punishment, 20 Crime &Just.: Rev. Res. 1 (1996); M. Matravers, Justice and Punishment: The Rationale of Coercion (2000).

first being filtered through the principles of justice. These OP-derived duties are grounded in the respect that all citizens should show to others' very capacity for and opportunity to citizenship. This, I should think, is clear in the case of taking another's life. But equally all instances of exercise of physical or psychological violence do not amount simply to depriving the other of a primary good. Nor are they akin to harms, which presuppose a basic acceptance of the terms of social cooperation. Take, for example, the trespassory prohibitions entailed by the institution of property. When one violates these prohibitions, one implicitly substitutes the holder of the property rights in their exercise. This presupposes either an acceptance of the grounds of property rights (e.g., one stealing another's property in order to enjoy it himself), or a rejection of the institutional scheme, which constitutes them (e.g., one could view property as unjust and therefore reject the very institution in practice). In either case, it presupposes the capacity of the holder of the property rights to be a participant in *some* institutional scheme, which will regulate access to and control over resources. With violence things are different, as it signifies an outright rejection of the other as a fellow participant in the social order. And only on the basis of respect for and mutual protection of the capacity of others to act as citizens is it possible to establish a fair political community. In other words, such natural duties provide the very conditions of possibility of citizenship and justice.

All this is not to suggest that natural duties do not derive from justice as well. According to Rawls, justice as fairness gives rise to the (natural) duty of justice, which "requires us to support and to comply with just institutions that exist and apply to us."[19] To the extent that we accept that the principles of justice as described by Rawls stem from the OP, then the duty of justice is of the same order as the duties I described, as it refers to the possibility of establishing and maintaining institutions of justice.

Natural duties are juxtaposed to *obligations* of justice, which are institution- and practice-bound and arise on the basis of voluntary acts. Obligations can arise only when relevant institutions have been set up, enabling citizens freely to pursue their conceptions of the good in coordination with one another.

Crucially, these natural duties and obligations set the only politically relevant measures of rightness and wrongness. An act is wrong only to the extent that it is a violation of a political duty set by the terms of the social contract.

19. Rawls, A Theory of Justice, supra note 5, at 115.

This already paints a different image than the, largely utilitarian, harm-centered understanding of wrongfulness. Harm here is incidental,[20] although this is not to say that it does not have some normative force, not in determining wrongfulness but in the process of deciding the appropriate response to a wrongful act. From now on, and because for my purposes the relevant similarity between them is that their violation is the political criterion of wrongfulness, I will refer collectively to natural duties and obligations as political duties.

Now, what does it mean for us to be under a duty, either positive or negative? Of course, it entails that we ought to act or abstain from acting in a certain way. At the same time, though, it entails that when we fail to act or abstain from acting in the required way, *we are to blame* for this failure. In other words, the wrong is attributable to us, we are *responsible* for it. Being responsible *for* something goes hand in hand with being responsible *to* someone. And this in turn entails that the one to whom the duty is owed and to whom the one owing the duty is accountable, has reason to criticize the latter and the standing (somehow) to hold the other accountable. Let me explain this by relying partly on Stephen Darwall's account of (moral as well as other types of) obligation in terms of the "second-person standpoint."[21] The second-person standpoint entails what Darwall terms a "circle of irreducibly second-personal concepts": *practical authority*, that is, the authority that A has over B if B has a *second-personal reason* stemming from a *valid claim or demand* held by A to whom B is *responsible* to the extent that all the rest holds.[22] This reveals something important about the structure of, at least certain, duties: being under a duty entails being responsible *for* its violation *to* whomever the duty is owed, with the latter acquiring a reason to criticize and the standing to hold the offender responsible for that violation.

I think there should be no doubt that duties derived from the OP are of the type of duty that Darwall has in mind. They are duties owed to others,

---

20. The understanding of legal wrongfulness in terms of the violation of some duties rather than harm is also present in Kant's doctrine of right. See A. Ripstein, Authority and Coercion, 32(1) Phil. & Pub. Aff. 35 (2004); and Force and Freedom: Kant's Legal and Political Philosophy (2009).

21. See S. Darwall, The Second-Person Standpoint: Morality, Respect and Accountability (2006).

22. Darwall's claim is very ambitious. He tries to ground the whole of morality on the second-personal standpoint, but this is not something on which I need to take a stance here. For his take on the law in terms of the second-person standpoint, see S. Darwall, Law and the Second-Person Standpoint, 40(3) Loyola of Los Angeles Law Review 891 (2008).

who can raise against us the claim or demand that we honor those duties. And in cases that we fail so to do, others have a reason to criticize and hold us responsible for it. The next question then is to whom we owe these OP-derived duties, and correspondingly, who has the standing to hold us accountable for our failure to honor them. For Rawls natural duties are owed to all as equally moral persons and irrespective of institutional relationships. This, and their unconditional nature, account for their "natural" property (and not, as I emphasized earlier, their being grounded in some prepolitical normative order).[23]

Crucially, I would argue that saying that they are owed to all means that they are owed both severally and jointly. This means, first, that no one can be arbitrarily excluded from those to whom the duty is owed. Secondly, it means that the duty is also owed to all as a collective, as a political community. Let us not forget that these duties are undertaken by the participants in the social contract precisely in their capacity as participants. This of course does not preclude a coincidence between the requirements of these political duties and the requirements of a moral order by which one feels bound. But the relevant reason for being bound by OP-derived duties is not the moral value that one attaches to them according to one's moral convictions, but that they are the product of a political agreement between rational and reasonable people, regarding each other as free and equal, and trying to establish fair terms of social cooperation.

When failing to honor those duties, then, we are equally accountable to the person who is directly at the receiving end of the violation (the victim, to put it simply), should there be one, as well as the community as a whole. I would suggest that this goes some way to justifying the authority of the state to organize responses to violations of political duties. This of course does not exclude the victim from the accountability-seeking process. But it does preclude the possibility of self-help and private punishment for the violation of political duties entrenched in the law.[24] It also sets the requirement of state involvement, to one degree or another, in procedures for the resolution of disputes between individuals. In other words, OP-derived duties do not lose their political character and become exclusively interpersonal as soon

23. Rawls, A Theory of Justice, supra note 5, at 115.

24. It does not necessarily follow from this that the State enjoys the monopoly of *all* responses, including punishment, to violations of norms. Whether these are justified in nonpolitical contexts is a different matter, which I am not interested here. What *does* follow is that the State is the principal actor in punishment in cases of violation of the law.

as they are instantiated in one-to-one relations between citizens. This is also the case for justice-derived obligations. Although they are undertaken between individuals, they are grounded in the requirements of justice informing the relevant institutions. The parties may be directly affected and have a direct and special interest in these obligations being observed, but this does not cancel out the interest of the whole political community in the changes in the allocation of primary goods or resources, which come about as a result of these obligations.

So far, I have argued that the violation of OP-derived duties provides those to whom the duties are owed—that is, the political community and, should there be one, the immediate victim—reason to criticize the offender and to hold him accountable. This, however, tells us nothing yet about *how* this criticism ought to be expressed or *how* the offender ought to be held accountable.

The first place to look for an answer to this would be retributivism, because it generally displays the same respect for the person as the Rawlsian theory of justice. I would, however, argue that retributivism of all kinds is unavailable to Rawlsian liberalism. "Retributivism" is a catch-all term, so I will consider only two central variants. Let us call "moral retributivism" the argument that the reasons for punishing are inherent in the wrongful nature of the act. To the extent that this implies a moral or otherwise necessary difference between criminal and noncriminal wrongs, then it can already be discounted, because it quite obviously relies on a comprehensive moral doctrine, which is incongruous with public reason.

This presents us with a textual problem. Rawls clearly toys with this type of retributivism, when he says, however in passing, that "the purpose of the criminal law is to uphold *basic natural duties*, those which forbid us to injure other persons in life and limb, or to deprive them of their liberty and property, and punishments are to serve this end."[25] The contradiction is clear. The range of political duties deriving either directly from the OP or from justice-based institutions is homogeneous in that they are all of the same type and their only salient characteristic is their political nature. This is clearly inconsistent with reserving the criminal law as the right response to the violation of some duties (the natural ones) but not others, which suggests that there is something particular and OP- or justice-independent about these duties. So something has to give. I would suggest that this must be Rawls's

25. Rawls, A Theory of Justice, supra note 5, at 314 (emphasis added).

obiter dictum regarding the criminalization of violations of natural duties alone.

It is also possible for one to be a moral retributivist in a different way. One may argue that irrespective of what it is that makes an act wrong, the *reason for punishing* someone for a wrongful act is moral. For example, one may hold the view that punishment is the only way of restoring the moral balance in the community. Once again, it is clear that such a substantively moral justification of punishment cannot be accommodated within a substantively *and* procedurally liberal state.

The second variant is what Alan Brudner has recently termed "legal retributivism" in one of the most sustained and interesting liberal accounts of crime and punishment.[26] Let me briefly outline the idea. Brudner argues that one deserves to be held accountable in a wide and rather open-ended range of cases of violations of others' rights, namely when one has: interfered with agency with the aim of demonstrating the nonexistence of rights; interfered with agency in the belief that this human being had no rights or belonged to a group whose members were right-less; knowingly interfered with agency or knowingly took the risk of doing so; performed actions short of actually interfering with agency but unambiguously manifesting an intention to interfere.[27]

However, the reason for *punishing*, rather than responding in some other way, is not located in the wrongfulness of these acts but rather in what Brudner calls, in a Kantian-Hegelian vein, the "vindication of the law's worldly authority":

> By *actualizing* a right-denying principle, however, the wrongdoer gave that principle an appearance of worldly authority, of existential force. He gave it, that is, the appearance of a law. That appearance would be allowed to stand if the nemesis of his principle were demonstrated by speech alone, for what is true in logic might not be true in the world; in the world, after all, there has been a violation expressing a claim to an absolute liberty—a claim that the civil remedy did not refute. So, by visiting the self-destructive consequence of the wrongdoer's principle upon him, punishment removes the appearance of its worldly validity and vindicates the worldly authority of Law.[28]

This view is indeed consistently liberal insofar as it draws reasons for the justification of punishment from the pedigree of the law and its claim to

26. A. Brudner, Punishment and Freedom: A Liberal Theory of Penal Justice (2009).
27. Id. at 41.
28. Id. at 47.

authority, and not from any normative order outwith it. To this extent, it is substantively impartial between moral beliefs. Nevertheless, it does not offer any procedural guarantees of impartiality. Notice how heavily the argument rests on a series of fundamental assumptions, which are largely controversial. For instance, for Brudner to be able to claim that some (or indeed, all) wrongful acts threaten the worldly authority of the law, it is presupposed that the maxim, on which one acts, necessarily takes the appearance of law. This, in turn, presupposes a strong view of reasons for action as agent-neutral and having the form of universalizable norms, thus excluding all other accounts of moral motivation. But this is precisely the type of substantive assumption that is excluded from public reason and therefore is out of place in a procedurally liberal justificatory scheme such as the Rawlsian one.

Such a substantively and procedurally liberal normative theory must rely on such a "legal" type of retributivism but without transgressing the limits set by the social contract, which is in turn constructed on the basis of the original foundational premises. Within such a liberal state, all justification must ultimately refer back to the terms of social cooperation. There is nothing in these terms and the corresponding political duties *logically* to necessitate specific responses to violations of these duties. This much is, I should think, obvious and uncontroversial. Similarly, as I have already argued, any moral connection between wrongdoing and response must be rejected. As far as I can see, the only remaining option is that the response to wrongdoing is justified not in terms of rightness or in any other necessary way, but rather in terms of the response's *appropriateness* for achieving an aim, which must be in line with and serve the terms of social cooperation. In the Rawlsian scheme this specific aim is dual: to maintain the long-term stability of the terms of social cooperation and to provide citizens with assurances of reciprocity, that is, that all will be sufficiently motivated to fulfil their political duties and respect the rights of others.

Viewing punishment as a matter of appropriateness though ultimately grounded in the justified response to OP-derived duty violation best captures, with some qualification, Rawls's distinction between justifying a practice on the one hand and instances falling under that practice on the other,[29] an argument that of course goes hand in hand with Hart's "hybrid" theory of punishment.[30] For Rawls and Hart, the institution of punishment may be

29. Rawls, supra note 14.

30. H.L.A. Hart, Punishment and Responsibility: Essays in the Philosophy of Law (1978).

justified on utilitarian grounds, but at the same time, the punishment of individuals must have retributive goals. Indeed, it would be inconsistent for any substantively and procedurally liberal theory not to reject both legal moralism and retributivism. Clearly, it would be unjustifiable for a liberal legal system to punish violations of moral, rather than political, duties. But it is equally clear that such a legal system cannot rely on a conception of punishment as necessitated by wrongdoing.

Many get rather wary at the thought of allowing consequentialist reason in the criminal law to such an extent, and strong objections have been raised against attempts at reconciling retribution and utilitarian goals. Nicola Lacey has forcefully and comprehensively voiced a number of such objections.[31] Viewing mixed theories of punishment as involving a move from act- to rule-utilitarianism, she argues that it is impossible to justify the set of criminal law rules in utilitarian terms and still prevent the principle of utility from determining how these rules ought to be applied in specific instances. Therefore, if the Rawlsian and Hartian hybrid theory remains unqualified, a clear-cut choice between either retribution or utility is inevitable. Weak retributivism—that is, the idea that the justification of punishment is primarily desert based, but the reasons for punishing are not conclusive—does not establish a duty on the state's part to punish. Whereas in the Rawlsian and Hartian hybrid accounts, considerations of retribution set the limits of utilitarian aims, here it is utility that constrains the scope of retribution. Lacey sees a tension between the weak retributive principle and the principle of utility because there is an overlap between their aims, yet an incompatibility between the ways they pursue them. Both purport to guide us about when punishment is *right*, but they must be mutually exclusive. And Lacey considers "a system in which the legislator made utilitarian generalization in framing the rules which were nevertheless primarily based on considerations" to be "highly dubious"[32] 2 as a prescriptive vision despite the fact that it may resonate with our actual practices.

These objections are fair, but I believe they can be rebutted if we disambiguate between the justifiability of responding to wrongdoing on the one hand and the appropriateness of punishment on the other. In this light, the answer is to be found somewhere in between the hybrid view of punishment and weak retributivism. The initial justification of punishment as a response

31. N. Lacey, State Punishment: Political Principles and Community Values (1988).

32. Id. at 55.

to wrongdoing comes down to the nature of political duties, which I discussed earlier. Whether this response will be punishment or not is to be judged in light of the causes of and possible solutions to partial compliance. It is indeed possible that there will be no need for punishment and that wrongdoing can be dealt with in alternative ways.

The fact that only in light of real conditions of partial compliance can the appropriateness of punishment be determined has, I think, been missed by writers, who have tried to reconstruct a Rawlsian theory of punishment.[33] The common locus of these attempts is the idea that, for punishment to be grounded in an impartial and fair manner, we must return to the OP but this time make more information available to the participants, namely the almost inevitable fact that ideal conditions will never obtain. The question then is, in light of the fact of partial compliance in real conditions, which arrangements for punishment will the participants agree on to make the distribution of penalties fair? I believe that the problem is located in the very question. To ask how *punishment* will be organized, participants must already know that punishment is entailed by or is the most appropriate response to at least some instances of wrongdoing. As I have argued, any logical or moral such links are unavailable for various reasons.

Now, for participants in the OP to know that punishment will be the most *appropriate* response to wrongdoing, they must have a great deal more information than the authors in question allow them. Not only must they be aware of the fact of partial compliance, but they must also know precisely the general motivational disposition as well as complex socio-psychological facts regarding the way that motivational deficit can be balanced out by the threat or actual implementation of punishment, information regarding the technologies of social control available to the community, and so on. The problem is that the more such information we include in the OP, the more the impartiality, which it is meant to guarantee, is watered down. The OP thus begins to resemble a Hobbesian state of nature in which self-interested beings try to strike a compromise and negotiate a modus vivendi that will prevent them from regressing into a nasty, brutish, and short life, than rational and reasonable

33. See S.J.M. Donnelly, The Goals of Criminal Punishment: A Rawlsian Theory (Ultimately Grounded in Multiple Views Concerned with Human Dignity), 41 Syracuse L. Rev. 741 (1990); S. Dolovich, Legitimate Punishment in Liberal Democracy, 7 Buff. Crim. L. Rev. 307 (2004); D.A. Hoekema, The Right to Punish and the Right to Be Punished, in John Rawls' Theory of Social Justice 329 (H.G. Blocker & E. H. Smith eds.) (1980). See also S.P. Garvey, Lifting the Veil of Ignorance 7 Buff. Crim. L. Rev. 443 (2004).

agents regarding each other as free and equal and trying to agree fair terms of social cooperation.

This does not mean that the justification of the criminal law and punishment is utilitarian in nature without any qualifications. First, the good that is pursued by punishment, that is, the aims of stability and assurance, is not arbitrary or a matter of some sort of felicific majoritarianism, for it is prescribed by and rests on justice. Therefore, whichever response is deemed the most effective will still have to be respectful of the status of citizens as free and equal participants in the political community, who have decided in common to which limitations they are willing to subject their freedom to act. These constraints of fairness do not derive from some notion of desert, some imaginary equivalence between the wrongdoing and the response, but rather refer to the procedures in which practices of punishment are to be decided. The utilitarian aims of assurance and stability alongside these constraints of fairness complete the sense of appropriateness of punishment. This clearly distinguishes between the fields of applicability of the principles of retribution and utility. It also rejects the mystical notion of desert without giving up on the idea of responsibility.

Some also understandably doubt that this consequentialism squares with the general orientation of the Rawlsian project. Thomas Pogge, for example, has argued that contractarianism regresses into a fully fledged consequentialism, and punishment is one context in which this is revealed, because the institutional arrangements that will flesh out the principles of justice are not visible from within the OP.[34] There are therefore no guarantees that, say, punishment will not be inflicted gratuitously and disproportionately.

Although, as I have already argued, Pogge is right about the conditions of institutional actualization being invisible in the OP, I do not share the worry that the Rawlsian scheme can regress into a consequentialist dystopia and thus cancel out its very point. This is because utilitarian reasons coexist with and are therefore tempered first by the requirement of responsibility and, secondly, by the requirement of publicity and democratic politics. None of these can by itself guarantee the reasonable use of punishment, but I believe that in their combination they can. I will argue this in two stages. I will start by outlining the conditions complementing the requirements of stability and assurance, and then tackle some specific problems in light of these conditions.

---

34. Thomas Pogge raises this challenge for the Rawlsian project in T. Pogge, Three Problems with Contractarian-Consequentialist Ways of Assessing Social Institutions, in The Just Society 241 (E. Franklen Paul, F.D. Miller Jr., & J. Paul eds.) (1995).

As I have been arguing, the criminal law's operative condition is the maintenance of stability and assurance. Let us understand punishment in the standard sense of the deprivation of sentenced offenders of some liberties and facilities, which they would ordinarily be afforded in their capacity as citizens. Is it a contradiction for a system of government geared toward achieving the reasonable maximization of everyone's liberty and fair social cooperation to curtail liberties in the name of stability and security? I would suggest there is no such contradiction to the extent that the curtailment of liberties meets four conditions: (a) It must be accepted by all as an eventual and meaningful response to the violation of responsibilities, which all have undertaken. It is therefore best understood as a voluntary, rational, and reasonable forfeiting of these liberties. (b) Inclusive, democratic political institutions and decision-making procedures must be in place through which such collective decisions will be made under conditions of publicity. (c) These institutions must be accompanied by institutions guaranteeing social justice as fairness in resource allocation and equality of opportunity.[35] This is the main way of securing stability and assurance. (d) Compelling, or as compelling as possible, empirical evidence must be available that, when all else fails, the threat of punishment is the most appropriate way of achieving stability. From the last two conditions it follows that the criminal law is to be used as an absolutely last *and* temporary resort, while taking measures to improve the basic structure with a view to achieving justice.[36] It also generally follows that

35. Retributivism of all kinds necessarily dissociates social justice from the criminal law, which it views as independent. Brudner's take on this is interesting. This is how he responds to the critical point that a retributive penal system cannot be just in the absence of social justice, guaranteeing that everyone will be on an equal footing by invoking the independence of the criminal law: "[O]ne can reasonably claim that an account of penal justice is exclusively valid for a just society only if social injustice makes private violence or takings permissible (as distinct from understandable) and therefore makes punishment for these actions unjust. Yet, apart from the necessity case, for which the model law itself prescribes, it does not. Social injustice can neither justify nor completely excuse (if moral permissibility grounds excuse) an individual's violating rights that the perfectly just society would itself protect; and so it cannot be a reason for no longer deserving the punishment for an action one would otherwise deserve." Brudner, supra note 26, at 324. A retributive penal order based on a conception of blameworthiness, independently of the fairness of the institutional structure of the political community, is an order that shows disregard for the liberty of citizens, for it treats them as subjects rather than as coauthors of their laws.

36. The argument that the criminal law should be reserved as a last resort is also made in D. Husak, The Criminal Law as Last Resort, 24 Oxford J. Legal Stud. 2, 207–35 (2004). See also R.A. Duff, A Criminal Law for Citizens, 14 Theoretical Criminology 293–309 (2010).

punishment, as we currently know and understand it, may not be an appropriate measure at all and should never be employed. This reveals the radical streak in the Rawlsian scheme of justice. It provides us with strong normative arguments for an institutional overhaul. It opens up a new perspective and allows us to view the criminal law not as a necessary institution and punishment not as a moral inevitability, but rather both as the continuation, rationalization, and humanization of historically contingent practices, which need either to be adjusted to serve the purposes of justice or to be jettisoned altogether.

The way has now been paved for addressing some specific issues, some of which have been raised by Pogge. Let us assume as a working hypothesis that all the sociological and socio-psychological evidence available to us indicates that the most effective way of preventing wrongdoing and therefore maximizing assurance and stability, at least in some cases and when everything else fails, is punishment. In that case:

1. What stops us from excessively criminalizing?
2. What stops us from punishing ad hoc with a view to the specific consequences of specific wrongful acts or the character of specific offenders?
3. What stops us from punishing less serious offenses (to the extent that we can hierarchize offenses in terms of their gravity) more harshly than more serious ones?
4. What stops us from punishing excessively harshly?

1. The Question of Excessive Criminalization

If no necessary link between offense and punishment is available; if part of the point of punishment is utilitarian, in the sense that part of what makes it an appropriate response to wrongdoing is the need to provide further motivational impetus and safeguard justice in the face of partial compliance; then why not wield the threat of punishment to preempt the effects of motivational deficit and, all the more so, when levels of compliance are evidently low? There will be nothing in the nature of wrongdoing or the content of the

---

The difference is that there the criminal law is seen as infringing rights that citizens would otherwise be afforded and allowed to enjoy, whereas I suggest it should be so because resorting to the criminal law in the first instance entails giving up on the possibility of citizens being motivated by their political and legal obligations.

violated norms necessitating or precluding punishment, therefore nothing able to set limits to punishment's scope. One may be punished for anything so long as one can be held responsible for his act. At the same time, I do not see how anyone, especially the liberal minded, would doubt that excessive criminalization is a rather unpalatable state of affairs. So, how can it be prevented despite its possibility?

Let us not forget two important points: first, that the primary aim is to establish fair terms of social cooperation; second, that the long-term stability of justice is to be safeguarded by a range of institutions but also, and crucially, by people's sense of justice, which will develop against the background of a just basic structure. Now, threatening with punishment for wrongdoing rests on the assumption that the more one has to lose, the more curtailed one's freedoms will be, then the stronger a disincentive he will have to violate community norms. It is also assumed that a sufficiently large number of people will display this motivational deficit. It follows then that overcriminalizing as a matter of general deterrence presupposes a general distrust in citizens' ability to abide by the law willingly out of the sense that they are coauthors of that law and out of a sense of respect for others, which requires everyone to respect generalized, normative expectations standardized through the medium of the law. In other words, an overcriminalizing state already regards its citizens as incapable of using their moral powers and therefore incapable of collectively agreeing, based on these moral powers, on fair terms of social cooperation. This, though, would already reveal a contradiction at the very heart of the state. It would be incoherent for a political community, which is supposed to realize the possibility of fair social cooperation, from the outset to assume a generalized failure on the part of citizens to be motivated by their status as parties in the social contract and a sense of justice. Therefore, criminalization can only be a reactive measure in light of actual instances of noncompliance and only when these instances are sufficiently widespread to warrant the threat of punishment. And, especially when the generalized failure to comply results from a justice deficit in the basic structure, the criminal law must be used as an absolutely last resort. Measures that do not impinge on people's liberty must be sought first.

The character of the criminal law as a last resort points to the solution to the second related question. Even if the threat of punishment is not to be used as a general deterrent in the fear of generalized noncompliance, what ought to be done in cases of *actual* generalized noncompliance? Imagine an anomic society, in which people show scant regard for the law, and all regularly and

systematically take the "bad man" perspective? I would suggest that in such cases the criminal law can do too little too late. In all probability such a society will suffer from basic structural problems, and it will have failed to engage all its members in inclusive, democratic, deliberative politics in light of public reason. This is not to argue simply that distributive injustice is criminogenic,[37] but that a generalized failure on the part of citizens to abide by the law will betray a failure of instituting law acceptable by all through democratic, inclusive procedures. The solution then is to be sought back on the drawing board rather than in the criminal law. A state that responds to noncompliance by way of coercion instead of by establishing the necessary conditions for social cooperation is simply not a just state, and it cannot become more just by punishing its citizens.

2. The Question of Ad Hoc Punishment

The rejection of desert qua direct and accurate translatability of offense into punishment and the argument that utility qua assurance and stability are central in the justification of punishment invite the familiar criticism that punishment can then be meted out ad hoc in view of the specific consequences of specific acts or, indeed, the actor's specific characteristics. Utilitarian theories of punishment are indeed vulnerable to this line of critique. However, and despite allowing room for some utilitarian considerations, the Rawlsian account is not affected because of the centrality of the fair distribution of liberties. As Rawls himself tells us, the distribution of punishment is subject to the requirements of the rule of law, which stem directly from the liberty principle.

The rule of law sets at least two constraints to punishment. First, no one may be punished absent a prior law. This is quite obvious in the case of institution-dependent duties. In the absence of a relevant institution, there is no duty to start with. But, as I have already shown, some wrongs stem directly from the OP, and are therefore preinstitutional but not prepolitical, some are justice-dependent and others are institutional. To the extent that liberty and equality are best served by a system of legal rules that is knowable in advance,

37. For the debate on whether there is a link between social inequality and crime, see indicatively: C. Grover, Crime and Inequality (2008); J. Hagan & R. Peterson eds., Crime and Inequality (1995); C. Coleman & J. Moynihan, Understanding Crime Data: Haunted by the Dark Figure (1996);); R. Reiner, Law and Order: An Honest Citizen's Guide to Crime and Control (2007).

it is not permissible to punish for acts that were not explicitly against the law at the time of their commission even though they may be preinstitutionally wrong. Of course, a legal system that fails to institutionalize and respond to the violation of OP- or justice-dependent duties will have failed to perform one of its central tasks. And citizens who do not realize (out of their sense of justice) their basic political duties fail in one of their tasks as citizens. But this does not give licence to the state to punish for the violation of duties that are not enshrined in law.

A necessary implication of this rule-of-law constraint is that the same penalties must be imposed for the same offenses. This follows directly from the requirement of a prior law that fixes the range of penalties and correlates them to offenses in a way that should preclude exceptions.

3. The Question of Proportionality

Once again, if there is nothing in the nature of an offense that can determine the type or severity of the appropriate response, then there seems to be no way of hierarchizing and differentiating between offenses in terms of their seriousness. Some of the potential results of this are again counterintuitive and possibly unpalatable. May a state really be at liberty to punish, say, violations of traffic legislation more harshly than causing others' death or violating others' sexual integrity based solely on utilitarian considerations of deterrence with the ultimate goal of maintaining assurance and stability?

For offenses to be ordinally scaled, there must be a standard against which they can be measured up. Obviously a political theory of criminal law and punishment cannot seek such a standard in the differential moral value of various goods. The whole point of social cooperation guided by public reason is that such substantive and essentially controversial judgments concerning value have no weight in public deliberation. However, this does not mean that there can be no *political* standards, that is, impartial standards acceptable by all, which can help us differentiate between offenses. The measure of seriousness will then be the extent to which the offense harms or threatens social cooperation. This brings us to the distinction drawn earlier in this paper. Some offenses threaten or undermine the very possibility of social cooperation. Without wanting to provide a comprehensive list, I would include here acts that turn against the bodily, psychological, and sexual integrity of others, and acts that discriminate and exclude the other from public life. Such offenses form a separate category of increased seriousness, with

their distinctive feature being that they undermine the very foundations of social cooperation. The remainder includes offenses that destabilize social cooperation by disappointing people's legitimate normative expectations without denying the very possibility of social cooperation. Within this category, offenses can be hierarchized in terms of the degree to which they destabilize the basic structure.

However, the hierarchization of offenses can still tell us nothing regarding the intensity of the response and, when that response is punishment, the scaling of penalties in relation to severity. To treat differentiated (according to seriousness) offenses in a differential manner, there must be a principle requiring that more serious offenses be punished more severely than the less serious ones. One way of extracting some such principle would be to resort to a notion of proportional desert. This, though, would commit one to a metaphysical or robustly moral view of wrongdoing and punishment, a thesis completely alien to a substantively and procedurally liberal theory of criminal law. In fact, any attempt at giving the proportionality principle a texture of necessity would commit us, in one way or another, to a strong metaphysical or moral doctrine.

This, I would argue, is a mistake committed by David Hoekema, who attempts to construct a Rawlsian theory of punishment by grounding the principles of allocation of penalties on the OP.[38] His argument is that, based on considerations of self-interest similar to the ones leading to the principles of justice, participants in the OP would opt for a retributive scheme of distribution of punishment, because the ordinal hierarchy that such a scheme guarantees would ensure that participants do not find themselves in the worst bad position once the veil of ignorance is lifted. This is complemented by a substantive normative claim. Hoekema argues that the disproportionality between the seriousness of an offense and the punishment made possible by purely utilitarian rationales of punishment violates citizens' self-determination. This latter thesis, however, begs the question to the extent that it presupposes that proportional punishment is a matter of respect for people's moral powers, a thesis that is not backed by Hoekema and, indeed, cannot be backed consistently with the non-moral orientation of the Rawlsian project. And if this argument is taken out of the equation, then Hoekema's thesis boils down to no more than an argument from stability: a just society will be at risk of being destabilized because people will not be sufficiently and appropriately assured

38. D. Hoekema, supra note 33.

that they will be able to pursue their idea of the good life, if they risk being disproportionately punished for offenses that are considered of lesser gravity.

Therefore, the only available solution is utilitarian in nature with deontological negative constraints. The relative severity of penalties will have to be determined in relation to the effects of offenses on the stability of social cooperation and on the sense of assurance of citizens. I said at the beginning of this section that relying on such a utilitarian argument for determining the intensity of punishment might lead to the unpalatable outcome of punishing more serious crimes with more severe penalties. And nothing I have said so far, including the possibility of distinguishing between offenses in terms of their gravity, seems to be able to divert us from that course. I would however suggest that it is precisely the unpalatability of this eventuality that will prevent it from happening. Although there is no way of ensuring that there will be some ordinal order, it is extremely improbable that people will not feel that more serious offenses call for more serious accountability-seeking measures, which will provide stronger guarantees of future deterrence, show due respect for the interests of the victim, and so on.

Now, if our worry is that criminal justice will be hijacked by certain groups with their own agendas, we should remember the three further conditions that the Rawlsian scheme sets: First, the rule of law, stemming directly from the liberty principle, requires that the law does not arbitrarily discriminate between people in the allocation of penalties, that offenders are punished only for the violation of preexisting laws, and so forth. Second, punishment is to be used strictly as a last resort. The first concern of justice is of course to set up a just basic structure, and the second to provide the conditions of stability for this just basic structure. Third, all norms are to be determined through democratic deliberation and decision-making and in light of public reason. This is a conception of democracy requiring that all be given the opportunity to participate in political decision-making, and it is therefore juxtaposed to a mere formalist majoritarianism. If these three conditions are successfully pursued, they should eliminate the possibility of the rule of law being used as pretence to persecute specific groups in pursuit of a political agenda.

4. The Question of Excessively Harsh Punishment

The last problem is related to the previous one, and so is its solution. The liberty principle and the rule of law impose formal constraints to the way that punishment is allocated, but this is where they meet their limits. They cannot

determine punishment substantively. From this and the consequentialist texture of the argument from stability and assurance, it follows that at first sight a state may inflict excessively harsh penalties even when rule-of-law requirements are met, the range of criminalization is narrow, offenders have fair warning, like cases are treated alike, and so forth. Indeed, even if some reasonable ordinal hierarchization of offenses and penalties has been achieved, there is nothing stopping us from raising the overall threshold of permissible punishment. Therefore, if it transpires that the deterrent and assurance effect of very long-term incarceration sentences for, say, driving offenses all but eliminate (it would be foolish to hope for complete elimination) such conduct, we would have a conclusive reason to impose such penalties, despite the intuitive disproportionality between offense and penalty. Now, I should think that there is no doubt that a state punishing with excessive severity across the board is an unpalatable state of affairs, and all our liberal intuitions tell us that this has no place in any well-ordered society that purports to be respectful of people's liberty. But how can such a state of affairs be averted in Rawlsian terms?

The answer runs along the same lines as that of the problem of ordinal proportionality. First of all, let us not forget that the point of a well-ordered society is the peaceful, productive, and fair coexistence of people with diverse life plans and conceptions of the good. In light of this, it is rather improbable that people would accept the possibility of a very severe punishment for what are considered to be minor offenses. They would be able to recognize and draw the boundary between deterrence or assurance on the one hand and complete incapacitation on the other. Part of the rationale of this is that most people would imagine themselves in a position of committing minor offenses. And we don't need to be behind the veil of ignorance for this to be the case. This is how a rational individual would think even in full knowledge of her personal circumstances. Still, generally law-abiding citizens would not be able to imagine themselves committing offences such as murder, rape, or robbery. Therefore, what is to guarantee that this law-abiding majority will not wish to incapacitate the minority more likely to commit more serious crimes (if such attempts at predicting patterns of the distribution of criminality hold any water at all)? The answer to this is to be found in the logic of the utilitarian calculus, which is much more complex than generally assumed. What must be taken into account is not only the general assurance and deterrence effect but also the effect a penalty will have for the offender, her environment, her ability to develop her life plan and pursue the good life despite her wrongdoing. This is because the Rawlsian scheme does not allow

the consequentialist calculation to lose sight of people's moral powers, which lie at the foundation of a well-ordered society in the first place. Although this does not mean that punishment is purely retributive, in the sense of just deserts, it is not purely utilitarian either in that it does not subordinate the interests and status of the few to the interests of the many.

Sure enough, there is still a remainder of risk here that the scheme could regress into a punitive dystopia.[39] For a procedurally liberal theory, this is not something that can be prevented from the point of view from nowhere. Nevertheless, sufficient constraints are in place to all but guarantee that things will not come to this. The view of the criminal law proposed here both is anchored to a sense of responsibility, albeit a political rather than a moral one, and places emphasis on the requirement of robust democratic deliberation and decision-making. Thus it shifts focus: instead of thinking that the criminal law can independently take care of justice, either by resorting to an a priori conception of blameworthiness and an equally a priori and perfectly corresponding punishment or to a calculation of utility from the external perspective and irrespective of responsibility, it views the question of punishment in a continuum with questions of justice and democracy, and gives priority to institutions guaranteeing fairness, including strong democratic institutions, subjecting the criminal law to these requirements. By doing so, it emphasizes the need for democratic government in conditions of fairness and conditions, and limits the coercive power of the state.

III. THE "DESERT ASYMMETRY" PROBLEM

Rawls's theory of justice is based on the rejection of a prejusticial sense of desert as a reason salient to the distribution of liberties and resources. Rawls is loath to allow accidental natural traits and advantages to determine the

39. Once again with the caveat that my intention is not fully to defend the Rawlsian view of the criminal law in this context but simply reconstruct it, I would suggest that the onus can be reversed. Is there *any* scheme that can eliminate this residual risk? It is hardly necessary to highlight this danger inherent in full-blown utilitarianism. Retributivism of all kinds, too, is not more reassuring. It places its hope in the existence of some exchange rate between wrong and punishment. But the inescapable, strong moral, and indeed mystical, undertone in this idea makes it all the more risky. Despite its adherence to the requirement of accountability, it offers far fewer, if any, guarantees that the question of punishment will not become prey to those with an interest in controlling the rest by way of punishment, but it also mystifies this in the guise of moral necessity.

allocation of social goods. What people deserve therefore coincides with what they are entitled to according to just social institutions. And yet, for Rawls desert seems to maintain its force as a criterion of the allocation of penalties. So, those who prove to be of "bad character" deserve to be punished for the wrong they have committed. There is clearly a potential inconsistency here. Rawls seems to pick the most convenient aspects of what it means to be morally deserving to fit his contractarian-consequentialist model, without allowing it to regress into a fully fledged utilitarianism when it comes to punishment.

Samuel Scheffler has tried to defend, rather than argue away, this asymmetry. He argues that according to "Liberal Theory," prejusticial desert is indeed impermissible as a ground for allocation of social advantages, but it can nevertheless serve as a reason for retributive punishment: "Whereas a just distributive scheme cannot coherently seek to reward moral desert, those whom the criminal justice system legitimately punishes have normally done something that would be wrong even in the absence of a law prohibiting it."[40] Scheffler justifies this differentiation in the function of the concept of moral desert in terms of the holistic character of distributive justice, which is juxtaposed to the individualistic nature of retribution. Whereas the former governs patterns of distribution from within a (just) basic structure, the latter is a "social response . . . to exercises of individual agency that society deems intolerable."[41]

I do not know that this resolves the asymmetry problem. First, it does not seem to be the case that distributive justice allocates goods and liberties exclusively in patterns, whereas punishment is necessarily individualistic. Entitlements to primary goods and basic liberties are, indeed, generally allocated according to some patterns set by the principles of justice. But *specific* individuals are entitled to *specific* goods on some conditions that relate *specifically* to them. To illustrate, everyone may be entitled to the primary good of health care by virtue of their moral powers and citizenship,[42] but for one to be entitled—or to "deserve," in a loose sense of the word—to register

40. Scheffler, supra note 15. Some powerful objections have been raised against Scheffler's attempt to defend such a differential attitude toward moral desert. See E. Mills, supra note 15; J. Moriarity, supra note 15. See also D. Husak, Holistic Retributivism, 88 Cal. L. Rev. 991 (2000). I am not revisiting these objections here because what I want to highlight about the argument is different.

41. Id., 978.

42. I am leaving to one side the question of who does or ought to count as a citizen.

with a *specific* medical practice, one must live in the area or meet some other such requirement.[43] Similarly, everyone is potentially subject to criminal sanctions by virtue of their being part of the political community. Moreover, punishment must be meted out according to general patterns so that no one may be exempted from the penal law as a result of the rule of law, which flows directly from the liberty principle. But a *specific* individual may be punished—or "deserves" to be punished, in a loose sense of the word—only to the extent that she has violated the relevant legal rule.

Nevertheless, the fact that both distribution and punishment require individuating, triggering conditions does not mean that they are identical or that they collapse into each other. There are two fundamental differences. First, whether the conditions of individuation are attributable to the agent or not is immaterial in the case of distribution. For example, one may have ended up living in a specific area and subsequently having to register with a specific medical practice quite accidentally. Or one may have intentionally moved to the area to be able to register with that practice, which she considers particularly good. What is of interest is the state of affairs forming the circumstances of the individuation of the allocation of the good, and not how these circumstances came about. For punishment, however, the triggering condition is exclusively the fact that the agent is responsible for violating a norm. Secondly, and more importantly, distributive justice refers to citizens' *entitlements*. Penal justice, on the other hand, refers to *responses* by the state to violations of norm. This entails that in the former case the state is under a duty to grant citizens basic liberties and goods, but it is not *necessarily* under a corresponding duty to punish.[44]

The problem then persists. Is it inconsistent to depend on responsibility, or "desert," in Rawlsian terms, the allocation of penalties but not the allocation of social advantages? I would suggest that there is indeed an inconsistency but only to the extent that we view responsibility qua desert in its standard, strong retributivist terms, which Scheffler does not rule out. The inconsistency lies not in that desert, as an individuating circumstance, plays a different part in each case, but rather in that in distribution of social advantages, desert is justice related, whereas in retributive punishment it becomes

43. All this on the assumption that free health care is a primary good and that there is an institutional structure that materializes this principle.

44. This echoes the distinction between positive and negative desert drawn in J.L.A. Garcia, Two Concepts of Desert, 5(2) Law & Phil. 219 (1986).

prepolitical. But in that case the solution to the asymmetry problem comes at a very high price, because it potentially undermines the neutral, impartial character of the justification of the basic structure and the terms of social cooperation. For one morally to *deserve* to be punished independently of an institutional framework, there must be a necessary link between wrongdoing and response. But if this is so, if we admit prepolitical duties, which we owe to each other with their violation justifying punishment without remainder, then it is difficult to see how we can stop short of allowing controversial moral doctrines to determine all the terms of social cooperation. Alternatively, even if wrongdoing is tied to the institutional framework, desert may imply that punishment is the a priori right response to wrongdoing. Once again, this allows a comprehensive moral doctrine in the whole scheme.

The problem can be solved by qualifying the concept of desert and understanding it as indicating no more than that one's wrongful actions are a necessary but not sufficient condition to punish her. In this minimalist sense of desert, one is *responsible* for her actions to those to whom the relevant duties were owed, and that the violation of these duties gives the latter reason to criticize A and take accountability-seeking measures. I believe that this reading of desert coheres with Rawls's dicta but also sits comfortably with the gist of Scheffler's argument: "The Liberal Theory, as I am interpreting it, does not offer an affirmative defense of retributivism against all these forms of skepticism. However, it agrees that wrongful conduct—that is conduct that would be wrong even if it were not illegal—is normally a necessary condition of just punishment."[45]

No sooner is desert in Rawls recast as responsibility for the violation of political duties than it becomes obvious that we do not need to trace its origins to a prepolitical sense of wrongfulness, which sits so uncomfortably with the Rawlsian scheme. As I showed in the previous section, accountability flows from the nature of the duties that we undertake in conditions of impartiality, with respect to the moral powers of all and with a view to drawing fair terms of social cooperation. Fair social cooperation provides the grounding of these duties as well as the wrongfulness of their violation.

The asymmetry problem is therefore explained away. There is no inconsistency between arguing that accidental features and advantages gained at birth should not determine distribution even while claiming that responsibility for violations of political duties may lead to punishment.

---

45. Scheffler, supra note 15, at 987.

CONCLUSION

This is the picture of criminal law and punishment emerging from the Rawlsian theory of justice. The fair terms of social cooperation include duties and obligations that are endorsable by all members of a society to the extent that the latter cannot produce public reasons, for which they ought not to be bound by these duties. It is part of the very idea of being under a duty that one is accountable *for* failing to honor that duty *to* those to whom the duty is owed. In turn, the latter have a claim against the former in cases of violations of these duties; they have reason to hold her accountable and take accountability-seeking measures. Nothing in this determines what these measures may be or their intensity. These are matters of *appropriateness* to be determined by the surrounding circumstances of the relationship between the parties, without forgetting that the reason for the response is the offender's accountability. In the case of political duties, the circumstances have to do primarily with the assurance of the participants in a well-ordered society that duties will be reciprocally honored and of the long-term stability of social cooperation. Criminalization and punishment are contingent, historically qualified means of achieving stability and assurance. Because punishment entails the curtailment of people's liberties, it is always secondary to measures guaranteeing the justice of the basic structure and the allocation of liberties, social advantages, and resources. In other words, punishment comes only as a last resort. At the same time, the institution of punishment is subject to the constraints of the rule of law, which stem directly from the liberty principle. However, when it comes to the questions of whether we can set upper limits to the intensity of punishment or whether penalties can be hierarchized in relation to the ordinal placement of offenses, no such constraints can be imposed. These questions depend on the proper workings of the utilitarian calculus in combination with the requirements of democratic norm determination and the respect due to all, even when entering utilitarian calculations.

As I said in the introduction, this paper's ambition was only to give an outline of a Rawlsian theory of criminal law and punishment without comparing it to other such theories. However, let me tentatively list some of the prima facie advantages of such a substantively and procedurally liberal theory of criminal law: (a) It is a purely political theory of punishment, which remains neutral between conceptions of the good steering clear of trying to impose and enforce any one of them. (b) It avoids the fallacy of strict retributivism, that is, the idea that not only do wrongdoers deserve punishment,

and only punishment, but also punishment of a specific type or intensity. (c) At the same time, it sets deontological constraints to utilitarianism, preventing the latter from riding roughshod and subjecting individuals to the welfare, or whims, of the majority. (d) In departing from the mystification surrounding strict retributivism and viewing punishment as a contingent institution, it has the force of providing the necessary theoretical bedrock and redirecting the enquiry into punishment toward an examination of the empirical conditions determining the latter.