Making *More* Sense of Retributivism: Desert as Responsibility and Proportionality[1]

J. ANGELO CORLETT

In 'Making Sense of Retributivism,'[2] I argued that John Rawls'[3] attempt to rescue utilitarianism from the threat of the possible punishment of innocents objection is unsuccessful in that it rests on his unsupported claim that utilitarian models of punishment provide plausible justification(s) for the institution of punishment, while retributivist models of punishment provide plausible justifications for particular forms of punishment. One of the main motivations of Rawls' distinction is to thereby provide utilitarianism with a deontological structure of rules that would somehow guard against the punishment of innocents for the sake of social utility maximization, a long-standing concern with act-consequentialist theories.

In response to Rawls, I argued that he mistakenly thinks that his distinction resolves the problem. On the contrary, his distinction only worsens it. For it would appear that a plausible model of punishment, whatever its philosophical orientation, must among other things provide *both* a moral justification for the institution of punishment *and* of particular forms of punishment. Otherwise, such models would be lacking in completeness and thereby run the risk of theoretical inadequacy. Thus the significance of Rawls' distinction for philosophical conceptions of punishment is not that retributivist and utilitarian models answer different questions about the justification of punishment. Rather, it is that any complete

[1] I sincerely appreciate the opportunity to respond to Thom Brooks' commentary on my article, as it presents me with an opportunity to clarify some points of philosophy and to thereby advance the dialectic between Brooks and me in a fruitful and collegial manner. I am also grateful to Anthony O'Hear for his helpful comments on how the manuscript could be improved.

[2] J. Angelo Corlett, 'Making Sense of Retributivism,' *Philosophy,* 76 (2001), pp. 77–110.

[3] John Rawls, 'Two Concepts of Rules,' *The Philosophical Review,* 64 (1955), pp. 3–13. Reprinted in John Rawls, *Collected Papers* (Cambridge: Harvard University Press, 1999), pp. 20–46.

doi:10.1017/S0031819103000287 ©2003 The Royal Institute of Philosophy

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

model of punishment must, among other things, provide plausible answers to questions of justification at each level.

In providing substance to my refutation of Rawls' point about different models of punishment and their respective normative attempts to justify punishment(s), I made brief reference to Immanuel Kant and some of his remarks about punishment in that Kant is more often than not construed as philosophically paradigmatic of retributivism, typically of the 'pure' (anti-consequentialist) variety. I extracted from Kant's work on philosophy of law a passage that begins to render problematic the traditionalist interpretation of Kant as a pure retributivist in an attempt to motivate philosophically my own version of retributivism. Indeed, Kant insists that the institution of punishment

> ... can never be used merely as a means to promote some other good for the criminal himself or for civil society, but instead must in all cases be imposed on him only on the grounds that he has committed a crime; for a human being may never be manipulated merely as a means to the purposes of something else. *He must first be found to be deserving of punishment before any consideration is given to the utility of his punishment for himself or for his fellow citizens*.[4]

This passage from Kant suggests pretty clearly that it is problematic to continue to attribute to him a retributivist model of punishment of the pure variety, whether or not it suffices to show that he is a mixed retributivist. In any case, it serves my more modest aim as a *starting point* for *my* own version of mixed retributivism, the plausibility of which would in the end serve to refute Rawls' implication that retributivism need not provide plausible answers to both questions of punishment.

In reply to my arguments, Thom Brooks[5] has made some critical claims about what I argued which are either false or misleading. In short, Brooks misinterprets my arguments as being aimed at interpreting the 'theory' of punishment of Kant (and even of Hegel!), and this affects his understanding of the general structure of my argument against Rawls and my attempt to provide, the *beginnings* of a new retributivist model of punishment. This misinterpretation

[4] Immanuel Kant, *The Metaphysical Elements of Justice,* John Ladd, Translator (London: The Macmillan Publishing Company, 1965), p. 100; *The Metaphysics of Morals*, p. 105, emphasis added. Quoted in Corlett, 'Making Sense of Retributivism,' p. 86.

[5] Thom Brooks, 'Corlett on Kant, Hegel, and Retribution,' *Philosophy,* 76 (2001), pp. 561–80.

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

on Brooks' part is ironic in that it is he himself who begins his commentary with the words of Kant: 'The duty of philosophy was ... to remove the deception arising from misinterpretation...'[6] Brooks misinterprets my basic argument and loses sight of its primary aim in my responding to Rawls on retributivist and utilitarian models of punishment in their respective attempts to justify punishment(s).

Although some of Kant's words on punishment serve as an *inspiration* for my own version of mixed retributivism, my claim that my view is 'Kantian' simply means that it draws significant inspiration from some of Kant's words on punishment. To think that I am involved in an 'interpretation' of Kant's 'theory' of punishment is to misconstrue my fundamental aim. That Brooks violates throughout his commentary on my critique of Rawls the interpretive principle of charity is perhaps most evident in the fact that he fails to even mention Rawls in his tome on Kant and Hegel on punishment. In effect, it constitutes a way of unwarrantedly shifting the direction of argument and analysis.

More specifically, Brooks suffers from the misconception that there is a theory of punishment in Kant, something that Jeffrie G. Murphy[7] and I have, respectively, gone to great lengths to dispute. It has been argued that a *theory* of punishment would need to address at least the following questions: What is the nature of crime and punishment? What are the moral and political justifications of punishments? What are the proper principles of criminal punishment? What are the appropriate punishments?[8] Although I concur with Murphy that there is probably no theory of punishment in Kant's writings,[9] and that there are five main questions that a theory of punishment must answer,[10] I nonetheless explicate what I think are Kantian answers to such questions, however tentative and sometimes problematic. In so doing, I believe that I provide some *foundations* of a Kantian theory of punishment. However, I stop well short of calling Kant's words on punishment a 'theory.' Nor am I interested in constructing his 'theory' from his words on punish-

[6] Brooks, 'Corlett on Kant, Hegel, and Retribution,' p. 561.

[7] Brooks is aware of Murphy's important article on the problem of attributing a theory of punishment to Kant, as he cites it at Brooks, 'Corlett on Kant, Hegel and Retribution,' p. 569.

[8] Jeffrie G. Murphy, 'Does Kant Have a Theory of Punishment?' *Columbia Law Review,* 87 (1987), pp. 509–32. Also see, J. Angelo Corlett, *Responsibility and Punishment* (Dordrecht: Kluwer Academic Publishers, 2001), Chapter 4.

[9] Corlett, *Responsibility and Punishment,* Chapter 4.

[10] Murphy, 'Does Kant Have a Theory of Punishment?'

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

ment.[11] If Murphy is correct, then we must be cautious in attributing to Kant such a theory. If I am correct, then perhaps the best we can do with Kant's words is to construct the very beginnings of a not altogether unproblematic view of retributive punishment.[12]

I shall leave it to Kant experts as to whether or not Brooks' attribution to Kant of the claim 'that governments *can* and *do* only impose *deterrent punishments*'[13] is correct. But even if it were accurate, as Kant is no stranger to making odd and unsubstantiated claims, it is unclear to me how it follows logically that 'any Kantian theory of punishment must be beholden to *deterrence*, not retribution, as the main justification for punishing.'[14] Besides violating the principle of interpretive charity by attributing to Kant an apparent is-ought fallacy in thinking that because governments can and do impose only deterrent punishments that deterrence is the main (moral) justification of punishing wrongdoers, this claim is problematic for another reason (unless, of course, one accepts the implausible claim that the issue is not one of normative justification of punishing wrongdoers). Assuming that Kant is as concerned as we are with the moral justification of punishing wrongdoers, then we can without conceptual absurdity ascribe to Kant *both* the claim that the state must be concerned with the deterrent effects of punishment *and* the claim that the moral justification of punishment raises a variety of other issues such as the unjustifiedness of punishing innocents, the limits of deterrence as a justification of punishing wrongdoers, etc.

More specifically, I understand Kant's above-cited words in the following way:

> ... Whether or not this makes Kant's view of punishment a 'mixed' one, the point is that it is no embarrassment even for Kant, or for retributivists in general, that appeals to social utility are made once the primary justification of punishment is provided. The primary justification of both the institution and practices of punishment is 'so that everyone will duly receive what his actions are worth. ..., And the fact that even on Kant's view of retribution non-desert factors are relevant to punishment indicates that it misses the point of retributivism to construe it as a view that holds that giving wrongdoers what they deserve as *all*

[11] See also J. Angelo Corlett, 'Foundations of a Kantian Theory of Punishment,' *The Southern Journal of Philosophy,* 31 (1993), pp. 263–84.

[12] Corlett, *Responsibility and Punishment,* Chapter 4.

[13] Brooks, 'Corlett on Kant, Hegel, and Retribution,' p. 566.

[14] Brooks, 'Corlett on Kant, Hegel, and Retribution,' p. 566.

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

> there is to justice. ... For retributivism *just is* the view which seeks and must seek a justification for both the institution and practices of punishment.[15]

Brooks interprets my construal of Kant's words otherwise: 'Corlett's Kant believes that we (a) affix criminality solely to retributive justifications and then (b) serve particular punishments to guilty persons due at least in part—to consequentialist concerns.'[16] However, there is a tremendous difference between what I write, above, and what Brooks takes me to mean by what I write. The fact that, on Kant's view, 'appeals to social utility are made once the primary justification of punishment is provided' in no way *requires* that, as Brooks interprets me to think of Kant, the state 'serve particular punishments to guilty persons due—at least in part—to consequentialist concerns.' On my understanding of the words of Kant that I cite, considerations of social utility *need not* figure into the moral justification of particular forms of punishment. The 'at least' locution in Brooks' formulation of my construal of Kant's words here seems to imply that consequentialist considerations must figure into a Kantian justification of particular forms of punishment, while my own formulation does not commit the retributivist to such a view. At each level of justifying punishment(s), considerations of social utility are sometimes relevant, but are in no sense necessary, to the justification of punishment(s). This preserves the mixed (or at least, impure) sense of retributivism that I believe is implied in the words of Kant to which I refer, above, without making Kant

[15] Corlett, 'Making Sense of Retributivism,' p. 87.

[16] Brooks, 'Corlett on Kant, Hegel and Retribution,' p. 568. See also page 573 where Brooks makes a similar kind of misattribution to me pertaining to Kant's cited (by me) words on punishment: 'Corlett therefore misinterprets Kant's entire penal project. Kant is not a retributivist at a primary level and then open to considerations of social utility at a secondary level, that of affixing specific punishments to particular criminals.' Once again, Brooks grossly misinterprets the nature of my project in refuting Rawls rather than explicating Kant's philosophy of punishment. But Brooks also reinterprets my construal of Kant's words. For it is not that Kant is a retributivist at the level of justifying the institution of punishment and a utilitarian or consequentialist at the level of justifying particular forms of punishment. Rather, it is that Kant is a retributivist at *both* levels of justification of punishment(s), casting doubt on what really counts as being a retributivist, namely, on the view that being a retributivist does not entail anything at any level of justification about the consequences of punishment. Contrary to Brooks, then, we have no good reason to think that I confuse 'these two standpoints when justifying' my 'own version of proportional punishment.'

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

into some kind of consequentialist wherein considerations of social utility are *necessarily* (and, perhaps, *primarily*) part of the calculation of what amounts to justified punishment(s).

I shall refrain from taking seriously Brooks' devotion of an entire section of his commentary to my simple remark (in a footnote) that 'a less sophisticated, yet nonetheless pure, version of retributivism than that articulated by Rawls is found in G. W. F. Hegel.'[17] I simply do not 'dismiss' Hegel's 'theory' of punishment, as Brooks alleges.[18]

Furthermore, Brooks attempts to undermine the punishment of innocents objection to utilitarianism of the act-consequentialist variety. Citing Murphy, Brooks argues that '...nearly all utilitarians are rule- or act-utilitarians in one guise or another. For nearly *all* utilitarians, as a rule, there is a prohibition against punishing innocent people.'[19] But there are problems here. First, Brooks' locution 'nearly' in the second sentence of the cited passage implies that there are utilitarians who do not explicitly prohibit the punishment of innocents for the maximization of social utility. Thus there is good reason, philosophically, to be concerned with some version of utilitarianism that would commit such a rights violation (The right in question is the moral right to not suffer undue punishment). So even Brooks himself implies that there is some version of utilitarianism that serves as a propadeutic for Rawls' devising of a rule-utilitarian model of punishment that would seek to guard against the punishment of innocents objection.

Secondly, even if no utilitarian in history ever held what Rawls and others construe as act-utilitarianism (as opposed to rule-utilitarianism), it would not follow that such a view is unworthy of philosophical consideration—as the content of Rawls' 'Two Concepts of Rules' demonstrates. For act-utilitarianism serves a dialectical purpose of inspiring a more sophisticated, more nuanced, version of utilitarianism: rule utilitarianism according to which certain rules would be devised, adopted and enforced to guard against the concerns of the punishment of innocents objection. One (perhaps, in the end, not insurmountable) problem with such a rule-utilitarian move is that it becomes difficult to see the genuine difference between rule-utilitarian and deontological standpoints concerning the matter of punishment. This is especially the case if the above-cited words from Kant indicate, as I argue, that he has for generations been misinterpreted as a pure

[17] Corlett, 'Making Sense of Retributivism,' p. 80, footnote 11.
[18] Brooks, 'Corlett on Kant, Hegel, and Retribution,' pp. 561, 580.
[19] Brooks, 'Corlett on Kant, Hegel, and Retribution,' p. 578.

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

(anti-consequentialist) retributivist when he might well be a mixed one, as I suggest. In other words, if deontological theories are not represented (or, possibly, *mis*represented) as never being concerned with consequences of our actions, then the burden of argument seems to be on the rule-utilitarian to distinguish rule-utilitarian views from deontological ones. That burden includes, among other things, an explanation as to how the rules adopted by such utilitarians do not simply collapse the model into deontologically-based retributivism, traditionally construed. Perhaps in the end what has been taken to be utilitarian and deontological models of ethics over the years amounts to a facile distinction between the views of such philosophers as Kant, on the one hand, and John Stuart Mill and Jeremy Bentham, on the other. If so, then we may continue to embark on a quest, insofar as we are interested in the history of ideas, for what these respective views amount to. Even so, the utilitarian cannot simply state rules to be adopted and enforced that are already held by retributivists[20] as a matter of principle without providing an account of how rule utilitarianism remains distinctly rule-utilitarian and does not collapse into what is essentially retributivism.

My own version of retributivism is *that theory of punishment that advocates the hard treatment by the state of an offender* (*through an institutionally approved system of due process*) *because the guilty offender deserves it, based on her degree of responsibility and in proportion to the harm caused by her wrongful act, omission, or attempt*. This definition of 'retributivism' is consistent with my discussion in my previous article. However, what I did not do in 'Making Sense of Retributivism' was to provide more than a modest discussion of the controversial notion of desert, the concept perhaps most associated with retributivism. I hope to satisfy this challenge to some meaningful extent now.

On my view, desert is not a primitive concept having merely intuitive force. Nor is it a basic belief, self-justifying and according to which all other retributivist concepts are justified. Rather, desert is cashed out in terms of both responsibility and proportionality. By

[20] As Rawls himself seems to do in his adopting the requirements of guilt and proportionality for punishment in articulating his brand of rule-utilitarianism. For while Jeremy Bentham spends a great deal of philosophical energy articulating utilitarian reasons for proportional punishment [Jeremy Bentham, *The Principles of Morals and Legislation* (New York; Hafner Press, 1948), pp. 178–88], Kant and retributivists in general have long held these two principles of punishment as *foundational* to their versions of retributivism.

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

'responsibility,' I mean a cluster concept including notions of guilt (by action, omission, or attempted action), fault, intentionality, voluntariness and knowledge in a blameworthy sense such that one is culpable of wrongdoing and liable to punishment and/or compensation. By 'proportionality,' I mean that the punishment and/or compensation one is forced to undergo and/or pay is commensurate with the harm(s) one caused to others. In other words, the punishment and/or compensation must 'fit' the crime or tort offense.

To say, then, that one strongly deserves to be punished and/or forced to compensate another is to say that she is responsible in the robust sense for some harm and that whatever punishment and/or compensation she is forced to endure must align well with the harm(s) for which she is responsible. This retributive notion of desert is *not* committed to the idea that any particular person *is* in fact sufficiently responsible such that she deserves to be punished and/or forced to compensate another. Indeed, it is not even committed to the position that *anyone* in fact satisfies the conditions necessary and sufficient for punishment and/or compensation. Rather, it holds that the conditions of responsibility in the liability and blameworthy senses must be satisfied in order for a criminal or tortfeasor to be punished and/or forced to compensate another in proportion to her wrongful conduct. A criminal who fails to act knowingly, for instance, fails to qualify for punishment on this account.

It must be borne in mind that the cluster of responsibility concepts admit of degrees of satisfaction by wrongdoers such that wrongdoers are either (broadly speaking) strongly, moderately, or weakly responsible for their wrongful conduct. To the extent that each condition is satisfied a wrongdoer is to be punished and/or forced to compensate another party in proportion to her wrongful harm of that party (or its designated estate, etc.). And mitigation accrues in cases where full responsibility is absent. Admittedly, there are likely to be relatively few cases, at best, that would qualify as ones for which a wrongdoer is fully responsible, meriting unmitigated punishment and/or compensation. Nonetheless, my version of retributivism leaves open the question of whether or not anyone in fact deserves full punishment, say, for a criminal act. This is an empirical question that must be answered on a case by case basis, all things considered. My purpose is to attempt to devise principles that would discover the plausibility of a retributivist analysis of punishment. It is no mark against retributivism even if it turns out that no one deserves punishment (due, perhaps, to genuinely excusing circumstances), or that no one deserves unmitigated punishment (due, perhaps, to genuinely mitigating circumstances).

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press

The point here is that the retributivist concept of desert need be no mysteriously primitive one, much less a self-justifying basic belief. Rather, it is none other than an admixture of two long-standing and essentially retributivist cluster concepts: responsibility and proportionality, as I have construed them. It is *this* notion of retributivism, then—not some straw person rendition of it—that must be refuted by its detractors. The absence of such refutation, however, seems to indicate (if I am correct both here and in 'Making Sense of Retributivism') that any plausible (positive) theory of punishment must make use of this concept of desert. For what positive theory of punishment would deny that only the guilty should be punished, and that only those responsible in the requisite sense be punished in proportion to the harm(s) for which the wrongdoer is being punished? And if it is true that these essentially retributivist concepts are required of any plausible positive theory of punishment, then it seems to follow that all such theories of punishment just are retributivist ones, at least, in a quite meaningful and significant sense. This, however, would caution Rawls from attempting to borrow without permission a concept central to retributivism in attempting to construct a conceptual alternative that, in the end, appears to collapse substantially right back into retributivism. Could it be true that, for all of the generations of brilliant philosophical discourse on punishment (much of which was devoted to refuting retributivism) that failure to understand the genuine essence of retributivism has now landed us in precisely the position from which so many incisive minds attempted to steer us away?

My critique of Rawls' attempt to rescue utilitarianism from the grips of the punishment of innocents objection is in no way contingent on what Kant or any other historical figure in philosophy (or any other discipline) says about punishment. I proffered a critique of Rawls' argument and articulated the beginning of what I take to be a novel version of retributivism that would perhaps evade some of the most important objections to retributivist conceptions of punishment. The content of my arguments along these lines ought not to be lost in any shift of the genuine focus of argument to matters of historical philosophy on which my arguments do not depend for their plausibility. My critique of Rawls' articulation of a version of rule utilitarianism on punishment attempts to make *more* sense of retributivism in the midst of many misunderstandings of its nature and aim.

San Diego State University

https://doi.org/10.1017/S0031819103000287 Published online by Cambridge University Press