

Kant on the Role of the Retributive Outlook in Moral and Political Life

Arthur Shuster

Abstract: Kant is regarded as one of the staunchest advocates of retributive punishment in the modern tradition. This essay makes the case that a careful reexamination of Kant's account of punishment is necessary, especially in light of liberalism's characteristic inability to give the powerful moral appeal of retribution its due. Kant attempted to provide a clear analysis of what we mean when we say that morality demands that punishment be "proportional" to the crime. According to Kant, punishment's retributive aspect—as distinguished from its deterrent or restorative effects—is primarily concerned with redeeming (negative) moral worth. This paper attempts to unpack this claim by examining Kant's discussions of judicial punishment, the conscience, and divine punishment, respectively. It concludes that as a result of serious unresolved difficulties in his arguments for retribution, Kant manages only to deepen the question of the morality of retribution rather than to give it a decisive answer.

Modern punishment theory has its roots in the natural right tradition that begins in the seventeenth century. According to this tradition, the state's use of coercion against its subjects is authorized by the subjects themselves, united in a body politic, in order to protect them against violent death and the unlawful dispossession of their property. This meant that punishment was justified primarily as a means of preventing crime. It was admitted that penal institutions could be designed to benefit society in other ways as well—for example, by helping to educate past offenders to live better, more productive lives—but only so long as these additional uses of punishment promoted a stable and prosperous civil society and did not undermine the effectiveness of deterrence. Retributive punishment—which is imposed retrospectively with a view to guilt—is not intended to provide any future benefit, and for this reason it was understood by the early architects of modern penology as irrational and cruel, and thus unjust. Accordingly, Hobbes's seventh law of nature is "*That in revenges* (that is, retribution of evil for evil) *men*

The author would like to thank Russell Muirhead and Thomas L. Pangle for commenting on earlier versions of this manuscript, and the anonymous reviewers for their helpful comments.

look not at the greatness of the evil past, but the greatness of the good to follow.”¹ Since retribution aims to pay back evil for evil “without respect to the example and profit to come,” it is nothing but “a triumph . . . tending to no end,” and thus contrary to reason as well as an act of “cruelty.”²

Kant was the first great modern thinker to react against this narrowing of the scope of punishment by way of excluding its retributive aim. While it is true that some recent scholars have challenged the long-standing view of Kant as a champion of retributive punishment, the debate is usually only about the extent to which he was willing to qualify his support for retribution in order to accommodate competing political concerns such as crime prevention, necessity, and honor.³ Thus it remains correct to look to Kant as a classic defender of retribution. Kant held that the meaning of retribution can be determined by looking to notions familiar to us from everyday usage. We say that criminals get what they truly deserve only when their punishment “fits” their crime in kind and degree, that their suffering ought to be “proportional” to their guilt, and that they should be made to pay back their “debt” to society. Kant believed, for example, that when a murderer is punished with death it is precisely the symmetry between his crime and punishment that morality, as distinguished from statesmanly prudence, requires. This aspect of the murderer’s punishment is to be understood as entirely independent of any of its further consequences, such as removing such a dangerous individual from civil society, deterring others from committing the same crime, or providing emotional satisfaction to the bereaved and indignant family of the victim.

The notions of “fit” and “proportionality,” “debt” and “desert” in punishment are recognizable to ordinary moral experience, but they are still far from being theoretically transparent. Accordingly, it was Kant’s aim—mostly in his *Doctrine of Right*, but partly in other works as well—to explicate these notions and to show how they were connected to practical principles of

¹Thomas Hobbes, *Leviathan*, ed. Edwin Curley (Indianapolis: Hackett, 1994), chap. 15, § 19.

²*Ibid.* Cf. Locke, *Second Treatise of Government*, § 8; Montesquieu, *Spirit of the Laws*, bk. 6, chap. 16; Beccaria, *On Crimes and Punishments*, chaps. 2 and 12; John Stuart Mill, *On Liberty* (Indianapolis: Hackett, 1978), 9.

³See Jean-Christophe Merle, *German Idealism and the Concept of Punishment* (New York: Cambridge University Press, 2009), chaps. 1–3; David Sussman, “Shame and Punishment in Kant’s *Doctrine of Right*,” *Philosophical Quarterly* 58 (2008): 231; Thomas Hill Jr., “Kant on Wrongdoing, Desert, and Punishment,” *Law and Philosophy* 18 (1999): 407–41; Mark Tunick, “Is Kant a Retributivist?,” *History of Political Thought* 17, no. 1 (1996): 60–78; Sharon Byrd, “Kant’s Theory of Punishment: Deterrence in its Threat, Retribution in its Execution,” *Law and Philosophy* 8 (1989): 151–200; and Don Scheid, “Kant’s Retributivism,” *Ethics* 93 (1983): 262–82.

reason, for it was the *rationality* of retributive punishment, according to him, that distinguished it from mindless revenge.

It is noteworthy that since Kant, many thinkers—from his contemporaries such as Hegel to later liberals such as John Rawls, H. L. A. Hart, and Joel Feinberg—have followed his example in arguing that there is a real need to restore to modern punishment its retributive dimension. Some of these later liberals who have written favorably about retribution have argued that retributive punishment is fully compatible with liberalism, and that (contrary to what early modern thinkers such as Hobbes believed) there is no real tension between the backward-looking concern with desert that characterizes retribution and liberals' forward-looking commitment to external liberty and security.⁴

How is it, then, that a recent friendly critic of liberalism—Stanley Brubaker—can still claim that liberals cannot justify retributive punishment on their principles or even explain its widespread appeal?⁵ Like Kant, Brubaker believes that proportionality is central to what we think makes punishment just, and that ordinary, practical knowledge of the appropriateness of retribution is as common in liberal societies as it is in any healthy political community. Brubaker suggests as one explanation of the preponderance in political life of the retributive outlook that retributive punishment performs an important expressive function: through it a political community affirms in emphatic fashion its shared conception of the human good and the rank-ordering of morally significant characteristics (virtues such as courage, vices such as injustice) by rewarding those individuals whom it admires and punishing those whose actions it condemns as immoral.

Consistent liberals, on the other hand, cannot justify retributive punishment because of their basic theoretical commitments. Most prominent strands of liberalism (Brubaker cites Lockean, neo-Kantian, and utilitarian versions in particular) insist on maintaining a strict neutrality toward the various conflicting views of the human good, as a result of their skepticism regarding that good. Constrained by this neutrality, liberals can impose “penalties” (distinguished by Brubaker from “punishment”) which are intended to deter crime and to compensate victims for their losses, since this may be accomplished without requiring individuals to obey the law

⁴See John Rawls, “Two Concepts of Rules,” *Philosophical Review* 64, no. 1 (1955): 3–32; Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Harvard University Press, 1999), 211–12, 276–77; H. L. A. Hart, “Prolegomenon to the Principles of Punishment,” in *Punishment and Responsibility* (New York: Oxford University Press, 1968), 1–27; Joel Feinberg, *Doing and Deserving* (Princeton: Princeton University Press, 1970), chap. 5. See also Herbert Morris’s famous essay “Persons and Punishment,” *Monist* 52 (1968): 475–501, and Richard Dagger, “Playing Fair with Punishment,” *Ethics* 103, no. 3 (1993): 473–88.

⁵Stanley C. Brubaker, “Can Liberals Punish?,” *American Political Science Review* 82 (1988): 821–36.

for what may be held by some to be the morally right reasons. But liberalism's neutrality toward the good denies the right to punish as a means of expressing moral condemnation of criminal acts, since doing so would seem to presuppose some standard of the human good.⁶ Thus, according to Brubaker, while liberals may deter, compensate, and rehabilitate, they cannot punish in the sense that most resonates with ordinary moral experience.

On Brubaker's account, when liberals try to make room for retribution in their penal schemes they do so either inconsistently or else by way of emptying the notion of retribution of its essential meaning. When societies punish, they surely do so for the sake of deterrence, but they also punish for the sake of retribution because they see it as a legitimate and independent *end* of punishment. On this basis, Brubaker criticizes Hart and Rawls, who deny that we ever punish *because* we want the guilty to "pay" for their actions. They claim that we only punish for the sake of deterrence, and that when we invoke the notion of just deserts we only mean to set an upper limit, or side-constraint, on public coercion to forestall punishment of the innocent and disproportionate punishment of the guilty. While such a limiting principle could well be reasonable and just, it would not, according to Brubaker, entirely satisfy our concern for retribution since, contrary to Hart and Rawls, we do in fact tend to believe that punishing the guilty is a moral end in itself.⁷ Since liberal societies do punish retributively (in Brubaker's sense), liberalism's failure to justify retribution creates a troubling dissonance between liberal theory and practice.⁸

Brubaker's analysis is provocative, but does his claim that liberals cannot punish apply to the classic attempt to reconcile liberalism and retributive punishment—namely, that of Kant? In this essay, I argue that the answer is yes and no. Yes, because Kant understood the moral appeal of retribution better than today's liberals do, and because he tried to defend retributive punishment in a form that corresponds much more closely to ordinary moral experience. No, because his final position contains serious unresolved difficulties. Nevertheless, I try to show that we can learn a great deal from Kant about what makes retribution so attractive (despite its sanguinary character) and about the prospects for justifying it on rational grounds—not least because Kant's sympathy towards the retributive outlook makes him more sensitive than recent writers to its most characteristic features.

It is true that Kant's understanding of human goodness as adherence to a categorical imperative that systematically excludes all empirical motives is initially somewhat difficult for us to bring to bear on a concrete practical question such as punishment, as Brubaker himself observes.⁹ Yet the

⁶Ibid., 821–22, 825.

⁷Ibid., 828–31. See also J. Angelo Corlett's criticism of Rawls in "Making Sense of Retributivism," *Philosophy* 76 (2001): 81–83.

⁸Brubaker, "Can Liberals Punish?," 833.

⁹Ibid., 826–27.

misconception that Kant's autonomous will is a listless being lacking all purpose is belied by the important place occupied by the concept of the highest good in his practical thought in general, and in his theory of punishment in particular. More problematic is the fact that Kant's commitment to external or negative liberty cannot be divorced from his commitment to noumenal or positive liberty, a feature of his thought that is responsible for a much-discussed tension in his view of punishment.¹⁰ I begin by examining Kant's teaching on legal or judicial punishment, and then go on to look at the role played by the notion of retribution in the inner life of the morally serious person, paying special attention to Kant's analysis of the conscience and to his account of the highest good.

Punishment and the State: Retribution as a Moral Duty to Others in Civil Society

According to Kant, the state of nature is a condition where each is completely free to act according to his own notions of what is right. Because such an anarchic state predictably leads to widespread violence, it is a condition devoid of justice. Therefore men must resolve to escape it by constituting civil society, in which the sovereign state is authorized to adjudicate all disputes among individuals regarding what is rightfully theirs.¹¹ Since the civil order is the sole bulwark against the lawless state of nature, and thus the fundamental condition of all right, preserving it from dissolution becomes the primary duty of the sovereign state. On this basis it would seem perfectly just, as well as prudent, for the state to discharge its duty by threatening would-be lawbreakers with force, and then carrying out those threats, on whose credibility the effectiveness of deterrence depends. The economy of deterrence would involve calibrating penalties to offset the advantages criminals stand to gain through crime, but not to be so terrible as to tempt benevolent judges and juries into applying them inconsistently (thus making punishment less

¹⁰On the challenges that positive liberty poses for liberalism, see Isaiah Berlin's famous essay "Two Concepts of Liberty," in *Four Essays on Liberty* (Oxford: Oxford University Press, 1990).

¹¹Immanuel Kant, *Metaphysics of Morals* (henceforward *MM*), trans. Mary Gregor (New York: Cambridge University Press, 1996), 6:305–13 (84–90). Whenever possible, the pagination of the Akademie edition of Kant's works will be given as the first set of page numbers (preceded by the volume number); the numbers in parentheses that follow refer to the corresponding pages in the cited English translation. Other works by Kant frequently cited in the text have been identified by the following abbreviations: *Critique of Judgment* (*CJ*), trans. Werner Pluhar (Indianapolis: Hackett, 1987); *Critique of Practical Reason* (*CPr*), trans. Lewis White Beck (New York: Liberal Arts Press, 1956); *Groundwork of the Metaphysics of Morals* (*Gr*), trans. H. J. Paton (New York: Harper and Row, 1956); and *Lectures on Ethics* (*LE*), trans. Louis Infield (Indianapolis: Hackett, 1980).

certain and less of a deterrent). In modern theories of the state prior to Kant's, the need to stave off the lawless state of nature was regarded as sufficient to justify such an essentially deterrent penal system.¹²

It thus comes as a great surprise to readers of the Doctrine of Right that although Kant agrees in most respects with Hobbes and Locke about the state of nature, in his thematic discussion of the right to punish (§49E) he denies that deterrence—or any other material benefit—can be the primary aim of punishment: “*Punishment by a court ... can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society.*”¹³ Not deterrence but retribution is the essential purpose of punishment, for “*only the law of retribution ... can specify definitely the quality and the quantity of punishment; all other principles are fluctuating and unsuited for a sentence of pure and strict justice.*”¹⁴ And what exactly is the principle of retribution, as dictated by strict justice? “*None other than the principle of equality (in the position of the needle on the scale of justice), to incline no more to one side than to the other ... whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself.*”¹⁵ Punishment “*must be inflicted upon [the criminal] only because he has committed a crime,*

for a man can never be treated merely as a means to the purposes of another or be put among the objects of rights to things: His innate personality protects him from this, even though he can be condemned to lose his civil personality. He must previously have been found *punishable* before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens. The principle of punishment is a categorical imperative.¹⁶

As the last quoted passage indicates, retributive punishment is made necessary, in Kant's view, by a certain requirement of morality. Even if the criminal were to lose his civil personality, and therewith his civil rights, he would still be immune from a punishment based on a purely consequentialist calculus since such a punishment would amount to treating him not as a person but as a thing, not as an end but as a mere means. How, then, does retributive punishment treat the criminal as an end? Before we go any further it would help to consider briefly why, according to Kant, human beings should be treated as ends in themselves.

¹²See, e.g., Hobbes, *Leviathan*, chap. 28 (beginning); Locke, *Second Treatise of Government*, §§ 7–13 and 87–88.

¹³*MM*, 6:331 (105).

¹⁴*Ibid.*, 6:332 (105–6).

¹⁵*Ibid.*, 6:332 (105).

¹⁶*Ibid.*, 6:331 (104–5).

Life, property, pleasure, in short all those things that contribute to our happiness, can be good things, but without dignity the goodness of these objects of our desires is substantially undermined. Kant argued that all such things are only conditionally good, depending on our having the sole absolute or unconditional good, which is inseparable from human dignity—namely, the morally good will, which performs duty for duty's sake.¹⁷ Since every human being possesses such a will, at least potentially, each of us must be treated as an ultimate end in itself. Accordingly, one of Kant's formulations of the categorical imperative—the supreme principle of morality—is “act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.”¹⁸ Treating someone as an end means acting toward him in such a way that our actions can be compatible with his choices and with his ends, regardless of the content of those ends. In other words, to treat men as ends is to respect their autonomous choices, regardless of whether their ends happen to be harmful or beneficial. One of the illustrations Kant gives of this formulation of the categorical imperative is that of making a false promise, which he says could never be understood as being compatible with the choice of the man whom we have deceived. Even if the fulfillment of our promise would result in general harm, we cannot make a promise knowing we would break it under such conditions without frustrating the choice of the person to whom we make our promise, or, indeed, without treating him as a mere means to our ends.¹⁹

In the section of the Doctrine of Right titled “Preliminary Concepts of the Metaphysics of Morals,” Kant defines a person as someone whose actions, as well as the *effects* of those actions, can be imputed to him as their author.²⁰ If one's actions fall short of what one can be constrained by law to do, they are morally culpable, and their “rightful effect” is punishment. Conversely, if one does more than what the law requires, this is meritorious, and the rightful effect of meritorious action is reward.²¹ Rewards and punishments, then, are imputable effects of our deeds just as much as any natural effects—we are the authors of our own punishments and rewards. If, for example, I assault someone, both the victim's injuries and my subsequent punishment can be imputed to me as their author. This is paradoxical, to be sure, since it is hard to imagine how a criminal could choose his own punishment. Yet as radical as this may sound, it is consistent with Kant's understanding of human action as necessarily having not only an empirical or phenomenal component, but also a moral or noumenal one. As Kant suggests

¹⁷*Gr*, 4:393 (61).

¹⁸*Ibid.*, 4:429 (96), second illustration.

¹⁹*Ibid.*, 4:429–30 (97).

²⁰*MM*, 6:223 (16).

²¹*Ibid.*, 6:227–28 (19–20).

by the section “Preliminary Concepts of the Metaphysics of Morals” as a whole, it is impossible to explain the everyday moral terms we use to describe human action—such as *person*, *deed*, and *crime*—without reference to concepts that have nothing at all to do with the mechanistically caused world described by natural science.²²

Kant does not deny that as a natural being having empirical incentives the criminal does *not* will to be punished. Yet when viewed as a *rational* being, the criminal necessarily wills his own punishment since he has willed the deed along with all of the effects it entailed.²³ On this basis, Kant argues that to treat a criminal as an end involves imputing to him, as well as carrying out, the rightful effects of his choice. In this sense, to punish a criminal would be to pay him the highest compliment, whereas to let him off the hook would be to condemn him as childish, irresponsible, and something less than a full human being.²⁴

According to Kant, only retribution satisfies the moral requirement of treating men as autonomous agents, since only it, as distinguished from consequentialist punishment, defines the imputable rightful effect of the criminal’s action. Consequentialist penalties have nothing to do with the criminal’s choice, since they look only to the future conditions of common safety or well-being. Retribution, on the other hand, establishes a proportionality, or likeness, in both quantity and quality—according to the principle of “like for like” or *lex talionis*—between the offender’s action (i.e., the crime) and its rightful effect (i.e., his punishment), and in this way retribution respects and fulfills the offender’s autonomous choice.²⁵ This talk of “proportionality” and “likeness” as that which truly expresses the criminal’s choice must strike the reader as rather vague, or even mystical, and at any rate not enough to pass the test of reason, which Kant himself purports to follow. We must remember, however, that Kant’s use of the language of physics and mathematics in his moral writings is *metaphorical*, intended by him to describe aspects of the moral world for which our ordinary language is inadequate,

²²Cf. Jennifer Uleman, “External Freedom in Kant’s *Rechtslehre*: Political, Metaphysical,” *Philosophy and Phenomenological Research* 68, no. 3 (2004): 585–91.

²³*MM*, 6:335 (108).

²⁴This view has also been expressed in Herbert Morris, “Persons and Punishment,” and in Jeffrie Murphy, “Marxism and Retribution,” *Philosophy and Public Affairs* 2, no. 3 (1973): 217–43.

²⁵In the context, Kant anticipates the objection that equal retribution “is not possible in terms of the letter” in every single case. No one can take an eye for an eye from a blind man. Kant argues, however, that it is possible to remain true to the spirit of the principle, if not always to its letter. In his example, Kant suggests that when an innocent person has been insulted, the punishment of the offender can satisfy the demand of retributive justice, even if it does not take the exact same form as the original insult, as long as the offender is made to feel shame in proportion to the outrage he caused his victim. See *MM*, 6:332–33 (105–6).

steeped as it is in our experience of empirical phenomena.²⁶ Nevertheless, we do ordinarily speak of proportionality in punishment as something that morality demands, and Kant wishes to suggest that this vague language actually points to an a priori moral world whose rational basis he believes himself to have uncovered through his critical philosophy (contained in his three *Critiques*).

It is for these reasons that Kant regarded retributive punishment as civil society's categorical moral duty to its criminals. Kant unambiguously illustrates the strictness of this obligation when he comments on the hypothetical example of a society about to dissolve by the consent of its members—when punishment can no longer have a deterrent effect. In such circumstances even the last murderer remaining in prison must still be executed “so that each has done to him what his deeds deserve.”²⁷

There is, however, a serious problem with the account I have provided so far, a problem that has led some Kant scholars to reject his view of punishment as incoherent.²⁸ The difficulty stems from the fact that morality and right are not identical for Kant, but rather the latter is understood to be contained within the former as a part of the whole. Moral laws are the unconditional dictates of pure practical reason taken together, which are, in turn, divided into two kinds, juridical and ethical, according to the manner of their incentives. Juridical laws, or laws of right, can have external incentives, meaning that conformity with these laws can be coerced, whereas ethical laws must serve as their own incentives, in the sense that conformity with them requires that we obey them for their own sake.²⁹ From the point of view of juridical legislation it does not matter what motivates us to act lawfully—whether it be a noble respect for the moral law or a base desire for profit. Thus, for example, I may fulfill a contract merely because I hope to engage in profitable commercial dealings in the future. Although it cannot be said that I have thereby acted ethically, I nevertheless have done all that is required of me by juridical law. For its part, the state can do everything it is authorized to do to constrain me to obey contract laws, but it has no right to inquire into my motivations for complying. But if this is the case, then if I should actually break the law, what business does the state have in punishing me for my “inner wickedness,” as Kant would have it? How can criminals be punished for their immoral motives if juridical legislation is indifferent to inner motivation?³⁰ There is also the related problem that, as Kant says elsewhere, while

²⁶MM, 6:232–33 (26).

²⁷Ibid., 6:333 (106–7). Cf. Kant's note at *Gr*, 4:430 (97), where he clearly implies that punishment is a strict duty to others.

²⁸See, e.g., Jeffrie Murphy, “Does Kant Have a Theory of Punishment?,” *Columbia Law Review* 87, no. 3 (1987): 509–33.

²⁹MM, 6:218–19 (20); cf. 6:214 (14).

³⁰One might object that in the Doctrine of Right retributive punishment is concerned not with the *ends* of a criminal's choices, but merely with their “form” as they relate to

God can see into our souls and know our true motivations, human beings have no such insight.³¹ How, then, could the state punish a criminal retributively “in proportion to his inner wickedness” if it cannot exactly know his true motivations?

I will address the second difficulty first. There is one passage in the Doctrine of Right that wrestles with the tension between the moral need to punish criminals as they deserve and the difficulty we have discerning one another’s true motivations, where Kant advances a tentative (and perhaps not completely satisfactory) solution. He tries to show that, in the case of capital crimes, the “fitting of punishment to the crime” will *always* occur by “imposing the death sentence in accordance with the strict law of retribution,” since “only by this is a sentence of death pronounced on every criminal in proportion to his *inner wickedness*.”³² To explain his point by way of illustration, Kant gives the example of two rebels: one who acts from the honorable (though mistaken) motive of wishing to depose a government he regards as illegitimate, and another who acts only for the sake of his private gain. Kant argues that both would receive what they deserve if they were sentenced to die. “Since the man of honor is undeniably less deserving of punishment than the other, both would be punished quite proportionately if all alike were sentenced to death; the man of honor would be punished mildly in terms of his sensibilities [i.e., valuing honor more than life] and the scoundrel severely in terms of his [i.e., valuing life more than honor].”³³ Thus, there would be no need for the state to inquire into the two criminals’ actual motivations; whatever their motivations happened to be, a sentence of death (as dictated by the principle of retribution) is the fitting punishment!

the choices of others, as Kant appears to state in the context of his discussion of the principle of reciprocity (*MM*, 6:230 [23–24]). But in that context, Kant seems to be speaking only about reciprocal relations between persons insofar as their actions, “as facts,” can have influence on each other—that is, wholly as external phenomena—and does not intend to say anything about imputation, a subject with which he had already dealt earlier. (Cf. translator Mary Gregor’s note c on p. 24.) It is hard to imagine what Kant could have in mind, if not the criminal’s immoral ends, when he speaks of his “inner wickedness.”

An issue to which I cannot give adequate treatment here is why retribution must follow some immoral actions but not others. The answer, I believe, has to do with Kant’s distinction between perfect and imperfect duties. Human beings are competent to punish transgressions of perfect duties, knowing that such transgressions are always morally culpable, whereas imperfect duties need not be violated just because our actions do not appear to conform to them.

³¹See Immanuel Kant, *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: St. Martin’s, 1965), A 551–52/B 579–80 (475).

³²*MM*, 6:333–35 (106–8).

³³*Ibid.*, 6:334 (107).

This would indeed be an ingenious solution to the problem were it to hold generally in all cases of possible crimes and motivations. Unfortunately, it is very unlikely that it *would* hold in all cases, and thus Kant's solution is partial at best. Nevertheless, what is important is that this passage shows that Kant refused to abandon his position that punishment must requite deeds in terms of what gives them their moral worth—namely, their motivations.³⁴ Perhaps Kant believed that, ultimately, we have to accept a certain amount of uncertainty in judging the motives of others, given the alternative. For Kant, not punishing for fear of making an error of judgment about motivations may have morally worse consequences than punishing and making the occasional mistake, since making such an error of judgment may at most cause undue and admittedly regrettable physical harm, while refusing to punish retributively altogether would amount to denying to all criminals their rationality, thus dehumanizing them.³⁵

In order to address the first difficulty—that juridical legislation's indifference to the incentive of action seems to rule out retribution—we must first specify more precisely what Kant meant by "right" as this concept is developed in the Doctrine of Right. Because of his status as an end in itself, man is the highest being in nature. In relation to the rest of nature man is an absolute master who possesses an unlimited right to use, transform, and even destroy nonrational beings according to his will.³⁶ In reality, however, there are many rational beings coexisting at the same time in a finite physical environment, each with an equally rightful claim to unlimited external freedom. This inevitably brings human beings into conflict over external objects—that is, over land, food, and other resources. As a result of the equality of all rational beings, men are required to observe the law of external reciprocity, according to which any action is right as long as it can coexist with everyone else's equal external freedom.³⁷ Reason authorizes enforcement of this law, since this can be done without impinging on anyone's freedom: any asymmetrical coercion of another is a hindrance to equal freedom, and thus the enforcement of the law of reciprocity can be justified "as a hindering of a hindrance to freedom."³⁸ This universal authorization of reciprocal coercion then becomes the source of all rightful or juridical—that is to say, external—legislation. In civil society, the entire authority to enforce the law of

³⁴I must therefore disagree with Herbert's and Fleischacker's attempts to interpret Kant's use of "desert" as signifying something that does not relate to moral motivations. See Gary Herbert, "Immanuel Kant: Punishment and the Political Preconditions of Moral Existence," *Interpretation* 23 (1995): 67–72; Samuel Fleischacker, "Kant's Theory of Punishment," in *Essays on Kant's Political Philosophy*, ed. Howard Williams (Cardiff: University of Wales Press, 1992), 202–5.

³⁵See also Herbert, "Immanuel Kant," 67–68.

³⁶Cf. *CJ*, 5:429–31 (317–19) with *MM*, 6:246–47 (40–41).

³⁷*MM*, 6:230–32 (24–26).

³⁸*Ibid.*, 6:231 (25).

reciprocal coercion belongs to the state, by virtue of which it can arbitrate conflicts and compel those who defy the authoritative arbitration to submit to it by force.³⁹

We must also keep in mind that while fulfilling our contracts and refraining from murdering others are juridical duties, which may be externally coerced, they are also at the same time moral laws, and as such they ought to be obeyed even without external compulsion. That is to say, duties of right are also necessarily duties of ethics.⁴⁰ This implies that when I break my contractual promise I not only disrupt the equilibrium of an external system of reciprocal rights but also act immorally and thus incur moral guilt. As we have already seen, the rightful consequence of moral culpability is punishment. We can now see that punishment in its retributive aspect is understood by Kant as *redeeming* moral guilt rather than as *enforcing* mere compliance with juridical laws (which is accomplished by deterrence). As such, it does not derive from the authorization to enforce the system of equal external freedom—from right—but rather from the categorical imperative which enjoins us to treat others as ends by respecting their autonomous choices—from the moral law in a more fundamental sense.

We can now see how Kant could have conceived of punishment as essentially retributive without contradicting his distinction between juridical and ethical lawgiving. In this we have a prime example of how, in Kant, morality not only coexists with, but also in certain cases qualifies and restricts, the administration of right. While the principle of right requires deterrence as a means of enforcing equal external freedom, the administration of punishment cannot be purely deterrent—deterrence cannot be its only justification—because that would violate the moral requirement that we treat human beings as ends. Deterrence can be a rightful aim of punishment, but only on the condition that it has met the requirements of the principle of retribution.

Such a reading of Kant as I have advanced would put him somewhat outside the boundaries of mainstream liberalism, according to which right is understood to be independent and sovereign in its own sphere and directed strictly to the preservation of external liberty. Some commentators have tried to bring Kant back in line with mainstream liberalism by denying the conclusion that retribution is a strict duty of the state in Kant's account. Most recently, Thomas Hill Jr. has claimed that Kant's "crucial thesis concerns our liability to suffer in the recognition of our own misdeeds, not our right or duty to make others suffer for theirs."⁴¹ Hill believes that, according to Kant, there is no duty, but only an authorization, to punish the guilty because his "retributive policies" are "not based on ... intrinsic desert; nor

³⁹See *ibid.*, 6:307–8 (86) and 6:312 (89–90).

⁴⁰*Ibid.*, 6:219–20 (21).

⁴¹Hill, "Kant on Wrongdoing, Desert, and Punishment," 409.

do they stand as fundamental moral requirements. . . . Rather, they are best understood as derivative features of a practice that requires independent justification."⁴² On Hill's reading, Kant's fundamental justification of punishment is deterrence, which Hill takes to be implicit in the state's function as the enforcer of civil society's laws. Those passages in the Doctrine of Right in which Kant appears to endorse a retributivist position of the sort I have laid out above should, according to Hill, be interpreted in light of the purely deterrence-based justification of the institution of punishment as a whole. Retribution, therefore, serves only as a necessary, but not a sufficient, condition of punishing the guilty; rather than being a strict duty it is only a restraint on the state's enforcement power keeping it from punishing the innocent, or from punishing the guilty disproportionately.⁴³ In short, there is no moral imperative to punish the guilty when no beneficial consequences can be expected to follow.⁴⁴

Mark Tunick has attempted to defend a similar reading of Kant.⁴⁵ He points to two passages in the Doctrine of Right where examples are given of justifiable exceptions to the law of retribution, which he reads as providing evidence that for Kant punishment is fundamentally consequentialist. In one passage Kant relates a hypothetical case in which a shipwrecked and drowning man saves himself by pushing another shipwrecked man off a plank that kept him from drowning.⁴⁶ Kant says about this case that a court should not punish the first man for murder, since "the punishment threatened by the law could not be greater than the loss of his own life" by drowning. Tunick interprets Kant to be implying that since there would be no point in having a penal law in this instance as it could have no deterrent effect, the purpose of penal laws, as such, must be deterrence.⁴⁷ Tunick attributes the same meaning to two more examples, in which Kant says that a soldier who kills his opponent in a duel and a mother who kills her illegitimate child ought to be exempted from capital punishment (as required by the principle of retribution) because in both cases the law against murder conflicts with an overriding incentive of honor (of a brave soldier in the one case, of a chaste woman in the other), albeit one created by a "barbarous and undeveloped" custom.⁴⁸

I believe that Hill's thesis goes too far. As I have tried to show in my close analysis of Kant's statements on punishment in the *Metaphysics of Morals*, his

⁴²Ibid., 429.

⁴³Ibid., 428–31. In his interpretation, Hill follows Sharon Byrd's "Kant's Theory of Punishment" and Don Scheid's "Kant's Retributivism." Byrd reads Kant's basic position as virtually identical with that of H. L. A. Hart (see Byrd, "Kant's Theory of Punishment," 183).

⁴⁴Cf. Hill, "Kant on Wrongdoing, Desert, and Punishment," 433–34, 438.

⁴⁵Tunick, "Is Kant a Retributivist?," 60–78.

⁴⁶Cf. *MM*, 6:235–36 (28).

⁴⁷Tunick, "Is Kant a Retributivist?," 64.

⁴⁸Ibid., 65–66. Cf. *MM*, 6:336–37 (108–9).

intention was to understand punishment as a moral category, which led him to try to connect it with the categorical imperative and to argue for retribution as its essential purpose. Thus, I do not think it is accurate to say that deterrence is the *only* justification of punishment and that retribution is not a duty to the criminal.

I do, on the other hand, think that Tunick's interpretation of Kant's exceptions to the law of retribution has some merit. At one point, Kant seems to want to suggest that the exceptions he mentions are not really dispensations from the law of retribution. For example, he says regarding the cases of the duelist and the mother who kills her illegitimate baby that these two individuals find themselves faced with two competing and equally binding imperatives—the honor of their station and the prohibition against murder—and that this clash of imperatives causes each to be thrown back into the state of nature, where killing does not count as murder.⁴⁹ Thus, according to Kant, retribution would not have to be exacted since there was no law against which to transgress in the first place. But Tunick's interpretation of these exceptions seems better than Kant's, on the latter's own terms. Kant's appeal to the state of nature to explain these exceptions seems rather contrived. Is the conflict between a just law (e.g., against murder) and a "barbaric" sense of honor really an insoluble quandary? It would seem less contrary to morality to require a man to defend his honor in court (or in the court of public opinion) than to allow him to take another's life unnecessarily. Similarly, although it might require a hard choice, it seems less contrary to morality to hold a woman responsible for not having a child she would be unwilling to raise than to allow her to kill that child with impunity once that child is already born. What Kant in fact seems to be doing here is trying to strike a statesmanlike balance between morality and competing political forces—namely, between the moral demand for retribution and the entrenched traditional sense of honor. The same seems to be true of Kant's discussion of the right of clemency, where his concern is clearly with stability of the civil order.⁵⁰

Thus, although I do not think that Tunick's ultimate conclusion—that, for Kant, the purpose of punishment is consequentialist—is warranted, I do think that he helpfully points to the fact that Kant failed to adhere consistently to his stated position regarding the categorical character of the duty to punish retributively. On the one hand, Kant wanted to affirm that retribution is a strict duty, while on the other he wanted to balance this moral imperative against other concerns, although it goes without saying that a moral duty cannot really be thought of as categorical if it is subject to justifiable

⁴⁹MM, 6:336 (109).

⁵⁰See *ibid.*, 6:334 (107–8), 6:337 (109–10). Cf. Tunick, "Is Kant a Retributivist?," 63–64.

exceptions.⁵¹ One can only speculate about how Kant would have responded had he been challenged to resolve this contradiction more consistently. Nevertheless, it seems to me that despite this inconsistency Kant does articulate a basis upon which we might try to justify retribution in the strong sense, understood as a duty rather than as a mere side-constraint, and as an independent end of punishment. We can then provisionally say in response to Brubaker's challenge, with which this essay began, that Kant (who was at least a certain *kind* of liberal) could punish retributively in the way Brubaker understands retribution.

We have now traced the arguments by which Kant expounds the meaning of retribution and defends it as the essential purpose of punishment in a moral civil society. The cogency of these arguments depends on whether Kant's claim that retribution preserves intact the inner dignity of the offender, as an autonomous being, is true. But, so far, we have only accepted this as a formulaic premise. What proof is there that the offender himself really does experience his punishment as dignifying and morally elevating? Kant's account of retribution in the *Metaphysics of Morals* is less than fully satisfying because it does not fully show how his formulaic pronouncements about retribution connect to our experience. The next section begins to address this question.

Conscience and the Retributive Outlook

What separates human beings from the rest of nature, according to Kant, is our capacity to act as beings subject to universal legislation. Since this capacity is innate to humanity, each may hold himself in the highest esteem and may claim a right to be respected by others on an equal footing, regardless of the position in which he has been placed by nature or society. Although Kant tends to emphasize this egalitarian basis of personal worth, he also acknowledges another source of worth for which we can be held in esteem—namely, the goodness of our actions. An example of the latter may be found in the familiar case of Kant's world-weary and reluctant

⁵¹There is another passage that is often cited by Kant scholars as showing him to be espousing what is essentially a deterrence theory of punishment. In his posthumously published *Lectures on Ethics*, Kant says that "all punishments imposed by sovereigns and governments are pragmatic; they are designed either to correct or to make an example" (*LE*, "Reward and Punishment," 55). While it is true that Kant is here saying that every punishment imposed by the state will necessarily be intended to correct or to make an example, he is not denying that such punishments may also meet the criteria of retribution—that is, by reflecting the crime as much as possible in kind and degree. Nevertheless, if in the final analysis there remains a real discrepancy between the two texts, more weight should be given to the *Metaphysics of Morals*, which was published in Kant's lifetime, as representing his mature view, than to his lectures, which were published without his supervision.

philanthropist who, “no longer moved by any inclination . . . tears himself out of this deadly insensibility and does the action [i.e., helping others in need] without any inclination for the sake of duty alone.” At that moment, “for the first time his action has its genuine moral worth.”⁵² Although, as a human being, the reluctant philanthropist already deserves respect, his difficult and rare self-conquest elevates still higher his standing in our eyes and in his own, and is responsible for the moral flourishing he experiences.

If innate humanity alone is not sufficient to account for the moral flourishing described in the reluctant philanthropist’s example, then how exactly can we account for it? Frequently, Kant argues that what underlies our admiration for good actions is really our respect for the moral law as it is manifested in and through good actions. But this alone does not exhaust the full meaning of the experience of moral flourishing, as Kant himself admits, and as brief reflection will show. In the first place, we know that contemplating our virtuous actions is positively enjoyable, and that this enjoyment lingers for a time after the virtuous actions have been completed—and thus presumably after the moral law has ceased actively manifesting itself through our actions.

Second, we believe ourselves to be entitled to this enjoyment of our virtuous accomplishments as part of our proper and fitting reward. As Kant puts it in the Doctrine of Virtue, “there is a subjective principle of ethical *reward*, that is, a susceptibility to being rewarded in accordance with laws of virtue: the reward, namely, of a moral pleasure that goes beyond mere contentment with oneself . . . and that is celebrated in the saying that, through consciousness of this pleasure, virtue is its own reward.”⁵³ Remarkably, Kant’s gloss on the saying that virtue is its own reward denies the literal interpretation of this saying. Virtue is *not* its own reward because virtue essentially involves a painful struggle against one’s natural inclinations.⁵⁴ Virtue does, however, make virtuous men entitled to further compensation, part of which is the moral pleasure of self-esteem. The case of self-reproach is similar to that of self-esteem. When we judge ourselves to have done wrong in some way, we believe ourselves to deserve the self-reproach, or pangs of conscience, that we suffer on account of our wrongdoing. Thus pride and self-reproach, both essential to the inner life of morality, may be regarded as self-imposed forms of reward and punishment.

In Kant’s view, the function of the conscience is analogous to that of a judge also in the sense that it makes the unjust man aware of his deserving of suffering in the form of judicial punishment. When the conscience passes judgment, it “pronounces the sentence of happiness or misery, as the moral results of the deed.”⁵⁵ The conscience does not merely identify certain types

⁵²*Gr*, 4:398–99 (66).

⁵³*MM*, 6:391 (154).

⁵⁴*Cf. ibid.*, 6:379–80 (145–46).

⁵⁵*Ibid.*, 6:439n (189n).

of actions as either morally good or bad. Rather, its judgment is also a “sentence” bearing consequences: it “either acquits or declares us guilty and deserving of punishment.”⁵⁶ It is not enough that self-reproach and repentance follow this judicial pronouncement of the conscience; the conscience is satisfied only “if [its judicial verdict] is felt and enforced.”⁵⁷ According to Kant:

The first effectual expression of this judicial verdict which has the force of law is moral repentance; the second, without which the sentence is inoperative, is action in accordance with the judicial verdict. If it does not result in *practical endeavor* to do what is demanded for the satisfaction of the moral law, the conscience is but an idle conscience, and however penitent we may be the penitence is vain so long as we do not *satisfy the debt we owe* to the moral law; for ... a *debt* is not satisfied by penitence, but by *payment*.⁵⁸

The “payment” of the “debt” we owe to the moral law, as distinguished from mere “penitence,” clearly refers to judicial punishment. Although Kant does not spell it out, his meaning seems to be that a person with a healthy conscience must be willing not only to accuse himself but to submit to punishment, knowing this to be the morally necessary consequence of his actions. If he does not, then he loses the basis for self-respect and is degraded in his own eyes. “Preachers must, therefore, impress upon their hearers that, whilst they must repent for their transgressions against their duties to themselves, though they cannot remedy these, *in the case of injustice done to others* mere repentance is not enough: it must be followed by endeavor to remedy the injustice.”⁵⁹ To salvage his dignity as a moral being the offender must submit himself to retributive punishment.

We can see in this account of the conscience that there is indeed some support in ordinary moral experience for Kant’s view of the dignifying effect of retributive punishment on the criminal himself.⁶⁰ That there is no parallel to this account in the works of Hobbes, Beccaria, or Bentham—the classic defenders of deterrence-based punishment—is a serious weakness of theirs, and a great credit to Kant. So far we have only dealt with the role of retribution as it applies to human punishment; we have not yet examined it in the context of divine justice, insofar as reason can tell us something

⁵⁶LE, “Conscience,” 129.

⁵⁷Ibid., 131.

⁵⁸Ibid., my emphases.

⁵⁹Ibid., my emphasis. As an illustration of what Kant may have in mind in this passage, consider the example of the titular character of Thomas Hardy’s *The Mayor of Casterbridge*.

⁶⁰This account bears an obvious resemblance to, and is likely to have been in some part influenced by, biblical moral psychology. On the conscience, see Romans 2:15 and Proverbs 20:27; on retribution, see Romans 2:5–12 and Proverbs 20:30.

about this. It will be helpful to investigate the latter as well, not only for the sake of completeness but also because, as we shall see in the next section, Kant's account of divine justice affords even deeper insight into the basis of the powerful moral appeal of retribution.

Divine Retribution and the *Summum Bonum*

According to Kant, when human beings are acting morally, they implicitly make certain theological assumptions or postulates, even when they are not aware of doing so. The postulation of the existence of an omnipotent, intelligent, and just being that always acts in accordance with the laws of morality arises out of a need to resolve the great "antinomy" of practical reason.⁶¹ As natural beings, we are driven by the desire to fulfill our needs, the total satisfaction of which we call happiness. On the other hand, as moral beings we hold that we ought to pursue virtue—a good higher than happiness and choiceworthy for its own sake—above all else regardless of whether or not it also contributes to our well-being. The highest good, or *summum bonum*, for human beings is to possess both goods at once, happiness and moral rectitude, even though there seems to be no necessary correlation between these two in nature. We do not think that by gratifying our desires we will become virtuous, nor do we have any reason to expect that by acting virtuously we will necessarily become happy. As both experience and natural science teach us, there is no necessary connection between acting well and faring well in the natural world.⁶²

Still, precisely in the moments when we are acting morally we cannot help but believe that the *summum bonum* is possible and, moreover, that we can bring it about through our own actions, for if it were otherwise then "the moral law . . . must be fantastic, directed to empty imaginary ends, and consequently inherently false."⁶³ From these facts of the active moral consciousness Kant deduces the general principle that if "virtue and happiness are thought of as necessarily combined, so that the one cannot be assumed by a practical reason without the other belonging to it," then this combination must be "as the connection of cause and effect."⁶⁴

The only way in which we can conceive of the connection between virtue and happiness as *causal* is if we postulate the immortality of the soul and the existence of a just and omnipotent God.⁶⁵ Such a God would guarantee the "effect" of happiness as a consequence of moral goodness by distributing well-being to each in exact proportion to his desert, "for to be in need of happiness and also worthy of it and yet not to partake of it could not be in

⁶¹*CPr*, 5:113–14 (117–18), 5:124–32 (128–36).

⁶²*Ibid.*, 5:110–14 (114–18).

⁶³*Ibid.*, 5:114 (118).

⁶⁴*Ibid.*, 5:113 (117).

⁶⁵*Ibid.*, 5:114–32 (118–36).

accordance with the complete volition of an omnipotent rational being."⁶⁶ Only on the assumption of such postulates is morality possible, since only then can we allow ourselves to hope that human beings will be rewarded and punished in accordance with their worthiness by an unerring judge, if not in this life then in the next.

This has the terrifying and awesome implication that each person will be held to account by an omniscient and all-powerful judge for every one of his actions. What is more, while Kant's God is a benevolent lover of mankind, he is above all a lover of justice for its own sake. For in distributing happiness in accordance with the concept of the *summum bonum*, God does not will that all men should necessarily be perfectly happy, but rather that they should enjoy as much happiness as they have deserved through their deeds. In the *Critique of Judgment*, Kant says that the final purpose of nature, as created by God, "can only be *man under moral laws*."⁶⁷ But he continues:

I say deliberately: *under moral laws*. The final purpose of creation is not man [acting] *in accordance with moral laws*, i.e., a man whose behavior conforms to them. . . . And this agrees perfectly with the judgment that human reason makes when it reflects morally on the course of the world. Even in evil we believe we perceive the traces of a wise reference to a purpose, provided we see that the wanton villain does not die until he has suffered the punishment he deserves for his misdeeds. . . . The highest wisdom in the government of the world we posit in this: that the opportunity for good conduct, but the consequence of both good and bad conduct, is ordained according to moral laws. In the latter consists, properly speaking, the glory of God, and hence it is not unfitting if theologians call it the ultimate purpose of creation.⁶⁸

According to Kant, then, the ultimate end of divine providence is not *necessarily* men acting in complete conformity with moral laws. Rather, the end of providence is perfect justice: the proportional distribution of happiness in accordance with moral worthiness.⁶⁹ In this, and not in God's unbounded benevolence, consists the "highest wisdom" of his governance.

This ringing endorsement of divine retribution is striking, and significant for our understanding of Kant's view not only of divine justice but of (human) judicial retribution as well. Kant had argued in the Doctrine of Right that human judges must embody the principle of retribution in order

⁶⁶Ibid., 5:110 (114–15).

⁶⁷CJ, 5:445 (334).

⁶⁸Ibid., 5:449n (338n).

⁶⁹See also *MM*, 6:488–90 (230–32); and "On the Miscarriage of All Philosophical Theodicies," in *Religion and Rational Theology*, ed. Allen Wood and George di Giovanni (Cambridge: Cambridge University Press, 2001), 8:260n (28n). Cf. Peter Byrne, *Kant on God* (Burlington, VT: Ashgate, 2007), 110–17; Lewis White Beck, *A Commentary on Kant's "Critique of Practical Reason"* (Chicago: University of Chicago Press, 1960), 270–71.

to protect human dignity by respecting criminals' rational autonomy. Yet, as we have just seen, in the context of the second and third *Critiques* we discover that God punishes retributively in order to bring about the correspondence of happiness and virtue. These seem to be two different accounts of the purpose of retribution. But are they essentially different? Kant's God is not the mysterious God of the Bible, who will be gracious to whom he will be gracious and whose ways are beyond our ken, but the perfect agent of practical reason—the same reason that guides ourselves when we act morally. Although the means through which God governs the world are hidden from us, the moral laws according to which he acts are perfectly known to us, since they are identical with our own. If we now recall our earlier discussion of what it means to respect a criminal's autonomous choice, we discover that (human) judicial retribution is not essentially different from divine retribution. Punishment, as we saw, was the rightful effect of morally culpable wrongdoing, which corresponds to the crime in proportion to the criminal's "inner wickedness." We also saw (in section 2) that injustice is a debt that must be repaid to the moral law, and that this debt is repaid in suffering that takes on a variety of forms, including punishment. It would not be misleading, then, to say that human retribution and divine retribution follow the same principle.⁷⁰

If one focuses on the Doctrine of Right alone it is easy to see how one might get the impression that law enforcement is the fundamental aim and justification of punishment as an institution, and that the principle of retribution is only a necessary, but not a sufficient, reason for punishing crime. If, however, one considers the Doctrine of Right in light of the second and third *Critiques*, as we have, it becomes much clearer that retributive justice—both in divine and human administration—is, for Kant, a core feature of the ultimate end of practical reason, the *summum bonum*. Taken together, this seems to be strong evidence for the thesis that Kant regarded retributive punishment of the guilty as an intrinsic moral good, and thus as a strict moral duty.

What can we now say about Kant's view of the morality of retribution, as it has come to sight? On the one hand, it seems to resonate with our own impression of one important aspect of justice.

When . . . someone who delights in annoying and vexing peace-loving folk receives at last a right good beating, it is certainly an ill, *but everyone approves of it and considers it as good in itself even if nothing further results from it*; nay, even he who gets the beating must acknowledge, in his reason, that justice has been done to him, because he sees the proportion between welfare and well-doing, which reason inevitably holds before him, here put into practice.⁷¹

⁷⁰This is Fleischacker's suggestion as well. See "Kant's Theory of Punishment," 203–6.

⁷¹*CPr*, 5:61 (63); my emphasis.

This is an easily recognizable expression of righteous indignation, which all or most of us have felt at some point in our lives. On the other hand, we are also compelled to wonder why God's (and our) sense of justice must be so strict and unyielding toward those who commit injustice. Would it not be better, from a moral point of view, if strict justice were occasionally softened by high-minded benevolence or by what Aristotle had called equity?⁷²

The most obvious, or most likely, Kantian response to this objection would be to repeat the steps of his derivation of the principle of the *summum bonum*: practical reason requires that happiness be dependent on virtue "as the connection between cause and effect," and this makes retribution a rational duty. But, as some of Kant's commentators have already pointed out, several problems with this connection between virtue and happiness make it difficult to accept. Lewis White Beck has argued that Kant does not adequately show how the *summum bonum* might be derived from the categorical imperative, given that the latter omits any reference to happiness.⁷³ Beck speculates that Kant's synthesis of virtue and happiness may yet be defended on the ground that "it is important for the architectonic purpose of reason in uniting under one Idea the two legislations of reason, the theoretical and the practical ... [since] reason cannot tolerate a chaos of ends." Ultimately, however, Beck himself does not believe that this explanation, based on the premise of "architectonic reason," has much force.⁷⁴

Susan Shell has observed that Kant's argument seems to blur together two distinct issues: *moral worth* as an unconditioned good, on the one hand, and *moral worthiness* as the condition of entitlement to happiness, on the other.⁷⁵ It makes sense that what is morally right should always be chosen over mere physical contentment if or when the two are in conflict. But this is not by itself enough to substantiate the claim that happiness should be distributed *in exact proportion* to virtue.⁷⁶ If happiness and virtue are both goods—albeit of unequal rank—why should the first be dependent on the second, and why should this dependence take the shape of mathematical proportionality? Why should we conceive of their relation as proportional to one another rather than cumulative? It is true that Kant maintains throughout his moral

⁷²In his own discussion of equity in the *Metaphysics of Morals*, Kant does not mention the possibility of dispensations from the law of retribution. According to Kant, claims of right based on equity are nonbinding because no judge can be appointed to render a decision. Cf. 6:234–35 (27).

⁷³Beck, *Commentary*, 242–45.

⁷⁴*Ibid.* Despite defending the *summum bonum* as a synthesis of "architectonic reason" Beck nevertheless denies that it has "any practical consequences."

⁷⁵See Susan Meld Shell, *The Rights of Reason* (Toronto: University of Toronto Press, 1980), 94.

⁷⁶Robert Taylor raises the same doubt in "Kant's Political Religion: The Transparency of Perpetual Peace and the Highest Good," *Review of Politics* 72 (2010): 11–12.

writings that happiness is a “conditional” good, but he seems to do so merely in order to emphasize that the good will is the only *unconditional* or absolute good. This alone is hardly sufficient to give Kant the right to conclude that the goodness of happiness ought to be conditioned *by virtue*.

That happiness ought to be dependent on virtue through the agency of God, as supreme dispenser of cosmic retribution, had been judged by Kant to follow from the immediate facts of the active moral consciousness, and the most relevant of these facts was said to be our hope that all men should attain just as much happiness as they deserve. But is the moral consciousness as unequivocal about this as Kant claims? Contrary to Kant, doesn't the moral consciousness itself approve of the softening or suspension of strict justice, under certain circumstances, on behalf of justice in a broader sense—as expressed, for example, by the principles of benevolence and equity? Is it not possible that the moral consciousness itself is much more divided than Kant admits about the goodness of strict retribution? Perhaps Kant is too quick to ascribe these concerns to “impurities” in our wills, rather than to justice itself.

Another possible defense against the difficulty we have raised may be constructed along the lines suggested by Emil Fackenheim.⁷⁷ Fackenheim argues that the key to understanding Kant's many contradictions lies in grasping his new kind of metaphysics. “Kant destroys the metaphysics which is based on speculation and replaces it with a metaphysics which is based on moral consciousness.” Kant “seeks to prove, not immortality and God, but that the belief in immortality and God is implicit in finite moral consciousness.”⁷⁸ We should therefore attribute the tensions and ambiguities we find in Kant's arguments—such as the ambiguity of the *summum bonum*—not to Kant's incompetent philosophizing, but to finite moral consciousness. “The philosopher too is a finite moral agent; and it is in his latter rather than his former capacity that he is in touch with ultimate moral reality: and as philosopher he recognizes this fact.”⁷⁹ Thus, according to this line of argument, Kant would urge us to embrace retribution, despite the unanswered questions we have raised regarding the relation between happiness and virtue, because it is a product of our finite moral consciousness.

But this explanation does not seem to me to escape the criticism I have been pressing. It is true that Kant argues forcefully for the independence of morality from theoretical science, but he never ceases to insist that, although independent of theoretical reason, morality is still grounded in (practical) reason. Without demanding any proof of the possibility of the correspondence of

⁷⁷See Emil L. Fackenheim, “Kant's Philosophy of Religion,” in *The God Within: Kant, Schelling, and Historicity*, ed. John Burbidge (Toronto: University of Toronto Press, 1996).

⁷⁸*Ibid.*, 9.

⁷⁹*Ibid.* Cf. 15–18.

happiness and virtue in the natural world, we may still inquire how reason is to resolve the contradiction in the moral consciousness between its uncompromising demand for exact retribution at one time, which Kant affirms, and its approval at other times of departures from strict retribution in favor of justice as benevolence and equity, which Kant for the most part ignores, but then reintroduces through the back door as exceptions to the *lex talionis*, without retreating from his claim that retributive punishment is a categorical imperative, as we have seen. This criticism implies not only that Kant should have refrained from characterizing retribution as a strict duty, but also, and more importantly, that the coherence of retribution as a whole becomes doubtful inasmuch as Kant has failed to supply answers for the reasonable questions we have raised and thus failed to fulfill his promise to justify retribution on rational grounds.

Conclusion

Working through Kant's reflections on the retributive outlook and its implications has deepened our appreciation for an aspect of punishment that transcends its utilitarian function as a mere law-enforcement mechanism. As we saw, Kant set out to articulate, clearly and systematically, what we really mean when we say that punishment ought to exact a "debt" owed by the criminal, or that punishment ought to "fit" the crime, or that we ought to treat criminals as responsible moral beings rather than as erring children. Kant paints a picture of the moral life as guided by a hope for the fulfillment of the *summum bonum* understood as the exact correspondence of happiness and virtue in the world, which he characterizes as perfect justice. This account of the moral consciousness and its desire for justice would imply that what punishment fundamentally means to us is not deterrence, or satisfaction of the victim's desire for vengeance, or even the declaration of society's disapproval of actions it judges to be intrinsically evil. Above all, to punish is to help give worldly existence to the moral conception of the dependence of happiness on virtue. The deep desire to see the virtuous flourish and the wicked punished shows its power by enduring in the soul despite the empirical evidence against the likelihood of its perfect realization, and by giving rise to the belief in an avenging God, as Kant so candidly admits. We owe Kant a debt of gratitude for his unflagging willingness to give theoretical expression to this desire. Through its enduring power in the soul the retributive outlook has defied classic liberalism's attempts to domesticate religious longings by continuing to nurture those longings.

Yet we are now in a position to say that Kant, too, has been frustrated in his attempt to give satisfactory justification to the desire for retribution, and so like most strains of liberalism Kantian liberalism cannot punish retributively, although Kant has perhaps gone further than any other liberal thinker in attempting to understand and defend retribution in its most robust form

(Hegel being the possible exception). As I have suggested, Kant had not paid sufficient attention to the ambivalence of the moral consciousness toward retribution and stopped short of taking the crucial step of subjecting to critical analysis the source of our hope for the exact correspondence between happiness and virtue. The latter sort of inquiry—perhaps in the form of a dialectical investigation of the opinions of the righteously indignant human type—would be needed in order to uncover the deepest meaning of this abiding human concern.