

# Kant's Theory of Punishment

THOM BROOKS  
*University of Sheffield*

The most widespread interpretation amongst contemporary theorists of Kant's theory of punishment is that it is retributivist. On the contrary, I will argue there are very different senses in which Kant discusses punishment. He endorses retribution for moral law transgressions and consequentialist considerations for positive law violations. When these standpoints are taken into consideration, Kant's theory of punishment is more coherent and unified than previously thought. This reading uncovers a new problem in Kant's theory of punishment. By assuming a potential offender's intentional disposition as Kant does without knowing it for certain, we further exacerbate the opportunity for misdiagnosis – although the assumption of individual criminal culpability may be all we can reasonably be expected to use. While this difficulty is not lost on Kant, it continues to remain with us today, making Kant's theory of punishment far more relevant than previously thought.

The duty of philosophy was, rather, to remove the deception arising from misrepresentation, even at the cost of destroying the most highly extolled and cherished delusion.

Immanuel Kant, *Critique of Pure Reason*<sup>1</sup>

Here, then, we see philosophy put in fact in a precarious position, which is to be firm even though there is nothing in heaven or on earth from which it depends or on which it is based.

Immanuel Kant, *Groundwork of the Metaphysics of Morals*<sup>2</sup>

## I. INTRODUCTION

Perhaps the most striking illustration of a retributivist position, in this case one that appears to be entirely backward-looking and oblivious to consequences, is found in ... Immanuel Kant.<sup>3</sup>

This may well be the most widespread interpretation of Kant's theory of punishment amongst contemporary theorists. Most commentators today would disagree with any suggestion that Kant's theory is open to consequentialist considerations.<sup>4</sup> On the contrary, I will demon-

<sup>1</sup> Immanuel Kant, *Critique of Pure Reason*, trans. W. S. Pluhar, Indianapolis, [1781/1787] 1996, p. 9 [A xiii].

<sup>2</sup> Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. and trans. M. J. Gregor, Cambridge, 1991, p. 35 [4:425]. When citing all works by Kant except the *Critique of Pure Reason*, the first number will refer to the page in the edition cited. In the brackets, the first number refers to the volume number of the Prussian Academy of Sciences edition of Kant's works with the second number after the colon referring to the page number in this volume where the quotation is found. Unless stated otherwise, all uses of emphasis in citations are given.

<sup>3</sup> Mark Tunick, 'Is Kant a Retributivist?' *History of Political Thought*, xvii (1996), p. 60.

<sup>4</sup> See H. B. Ashton, *Kant's Moral Philosophy*, London, 1970; Richard Brandt, *Ethical*

strate that Kant's theory of punishment is not purely retributive.<sup>5</sup> I believe that much of the reason for interpretative confusion is due to the very different senses in which Kant speaks of punishment regarding moral and positive law. When these standpoints are taken into consideration, I believe Kant's theory of punishment is more coherent and unified than previously thought.

I will argue that Kant's writings often justify punishments for transgressions of *positive* law on non-retributive grounds. This interpretation is contrary to most of the scholarship in the field. The confusion lies for the most part with Kant's retributive concerns regarding violations of the moral law, in my view. Most commentators seem to understand the plentiful discussions of punishments for transgressing moral law as either Kant's entire theory of punishment

*Theory*, Englewood Cliffs, 1959, pp. 496–8; Richard Dean, 'Cummiskey's Kantian Consequentialism', *Utilitas*, xii (2000); Samuel Fleischacker, 'Kant's Theory of Punishment', *Essays on Kant's Political Philosophy*, ed. H. Williams, Cardiff, 1992; Katrin Flikschuh, 'On Kant's Rechtslehre', *European Journal of Philosophy*, v (1997); Mary J. Gregor, *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*, Oxford, 1963; Jean Hampton, *Political Philosophy*, Boulder, 1997, pp. 134 f., 147, 175; Sarah Williams Holtman, 'Toward Social Reform: Kant's Penal Theory Reinterpreted', *Utilitas*, ix (1997); Conrad D. Johnson, 'The Authority of the Moral Agent', *Consequentialism and its Critics*, ed. S. Scheffler, Oxford, 1988, p. 273; Christine M. Korsgaard, *Creating the Kingdom of Ends*, Cambridge, 1996; Jeffrie G. Murphy, 'Kant's Theory of Criminal Punishment', *Proceedings of the Third International Kant Congress*, ed. L. W. Beck, Dordrecht, 1972, p. 434; Edmund L. Pincoffs, *The Rationale of Legal Punishment*, Atlantic Highlands, 1966, p. 9; Louis P. Pojman, 'For the Death Penalty', *The Death Penalty: For and Against*, ed. L. P. Pojman and J. Reiman, Oxford, 1998, pp. 1, 7–11; John Rawls, *Political Liberalism*, New York, 1993; John Rawls, *A Theory of Justice*, Cambridge, 1971; Jeffrey Reiman, 'Why the Death Penalty Should Be Abolished', *The Death Penalty*, ed. Pojman and Reiman, pp. 90–2; Patrick Riley, *Kant's Political Philosophy*, Totowa, 1983; Hans Saner *Kant's Political Thought: Its Origins and Development*, trans. H. B. Ashton, Chicago, 1967; Nancy Sherman, *Making a Necessity of Virtue: Aristotle and Kant on Virtue*, Cambridge, 1997; Avital Simhony, 'Was T. H. Green a Utilitarian?' *Utilitas*, vii (1995); Michael Slote, *From Morality to Virtue*, Oxford, 1992; Roger J. Sullivan, *Immanuel Kant's Moral Theory*, Cambridge, 1989, pp. 243 f., 361n25; and Howard Williams, *Kant's Political Philosophy*, Oxford, 1983.

<sup>5</sup> See my 'Corlett on Kant, Hegel, and Retribution', *Philosophy*, lxxvi (2001); B. Sharon Byrd, 'Kant's Theory of Punishment: Deterrence in its Threat, Retribution in its Execution', *Law and Philosophy*, viii (1989); J. Angelo Corlett, 'Making Sense of Retributivism', *Philosophy*, lxxvi (2001); David Cummiskey, *Kantian Consequentialism*, Oxford, 1996; David Cummiskey, 'Kantian Consequentialism', *Ethics*, l (1990); Robert P. George, *Making Men Moral: Civil Liberties and Public Morality*, Oxford, 1993, pp. 148–53, esp. 149n39; R. M. Hare, 'Could Kant Have Been a Utilitarian?' *Utilitas*, v (1993); Don Scheid, 'Kant's Retributivism', *Ethics*, xciii (1983); and Tunick, 'Is Kant a Retributivist?' Also see A. Phillips Griffiths, 'Kant's Psychological Hedonism', *Philosophy*, lxxvi (1991); Frederick Neuhausser's editorial footnote at p. 245n14 in Johann Gottlieb Fichte, *Foundations of Natural Right*, ed. F. Neuhausser and trans. M. Baur, Cambridge, [1795–1796] 2000; and David Weinstein, 'Between Kantianism and Consequentialism in T. H. Green's Moral Philosophy', *Political Studies*, xli (1993), esp. p. 631 where he claims that Green reinterpreted Kant as a consequentialist.

or some ideal penal theory divorced from empirical concerns.<sup>6</sup> In addition, some interpreters note the different senses in which Kant discusses punishment, but fail to see his theory of punishment as coherent.<sup>7</sup>

My reading uncovers a new problem in Kant's theory of punishment: How are we to assess the intentional attitudes of criminals in attributing blame and possible sentencing? The potential difficulty lies in his pronouncement that we can never know for sure the criminal culpability of any potential offender, making the possibility of our punishing innocent people a troubling reality. By assuming a potential offender's intentional disposition without knowing it for certain, we further exacerbate the opportunity for misdiagnosis – although the assumption of individual criminal culpability may be all we can reasonably be expected to use. While this difficulty is not lost on Kant, it continues to remain with us today, making Kant's theory of punishment far more relevant than previously thought.

## II. IS KANT'S THEORY OF PUNISHMENT RETRIBUTIVE?

For Kant, '[t]he law of punishment is a categorical imperative'.<sup>8</sup> Being a categorical imperative, the law of punishment must hold universally with necessity for all human beings, leaving no one outside the law's grasp.<sup>9</sup> Moreover, there should not be any contradictions amongst satisfactory legal maxims. If any such contradictions arise, then the maxims leading to them are unsatisfactory. Thus, laws of punishment are by extension universal laws that do not allow for any special cases.

The categorical imperative is also known as 'the moral law'. Under

<sup>6</sup> An interesting exception is Williams who claims that Kantian punishments must be justified from two standpoints: (1) the moral law's retributivism and (2) an empirico-utilitarianism. Therefore, punishments ought to both fit crimes and deter potential offenders. (Williams, p. 106; see p. 101.) In addition to Williams, Fleischacker confuses these two standpoints (Fleischacker, p. 193).

<sup>7</sup> See Jeffrie G. Murphy, 'Does Kant Have a Theory of Punishment?' *Columbia Law Review*, lxxxvii (1987). At 509, Murphy says: 'I am not even sure that Kant develops anything that deserves to be called a *theory* of punishment at all. I genuinely wonder if he has done much more than leave us with a random (and not entirely consistent) set of *remarks*'. Holtman attempts to reinvent Kant so she can 'reject the law of retribution and other disturbing details, but embrace more basic aspects of the accounts of justice and punishment'. Nevertheless, she says that his theory is at best incomplete. (Holtman, 12, 16, 21; also see Hannah Arendt, *Lectures on Kant's Political Philosophy*, ed. R. Beiner, Chicago, 1982, pp. 7 f.; Stuart Brown, 'Has Kant a Philosophy of Law?', *Philosophical Review*, lxxi (1962); Fleischacker, p. 193; Saner, p. 2; and Slote, pp. 31–57.)

<sup>8</sup> Kant, *The Metaphysics of Morals*, ed. and trans. M. J. Gregor, Cambridge, 1996, p. 105 [6:331].

<sup>9</sup> Necessity arises from a universal duty to respect the moral law. (Kant, *Groundwork*, p. 13 [4:400]. See *ibid.*, p. 14n [4:402].)

the moral law, all laws are to be determined independently of our desires and inclinations in an effort to ensure universality.<sup>10</sup> This rejection of a subjective standpoint has led most of Kant's critics to charge his ethical theory with empty formalism, mainly for its having the appearance of being an ethics based purely on logical criteria.<sup>11</sup>

For Kant, when we perform a particular action, it is supposed that we are stating to our community that all may act in the same manner as we have.<sup>12</sup> As an example, if a thief stole property from someone, the thief's action would effectively establish a universal maxim that everyone might do the same. Likewise, if a murderer were to kill an innocent victim, the murderer's action would by extension establish a universal maxim sanctioning the ability of any citizen to kill innocent victims. In addition, criminals use other persons as a means to their own particular ends, in making themselves an exception to universal principles. This is to say that a thief desires to steal property so that she might benefit from its ownership: The thief would not like her belongings stolen from her by a second thief. For Kant, the use of other human beings as a means to an individual's subjective end for any reason is morally objectionable.<sup>13</sup> The difficulty with theft and murder – amongst many other crimes – is that they *ought not* to become

<sup>10</sup> See J. Mariña, 'Kant's Derivation of the Formula of the Categorical Imperative: How to Get it Right', *Kant-Studien*, lxxxix (1998), pp. 167 f.

<sup>11</sup> Kant says: 'All I need for morality is that freedom does not contradict itself and hence can at least be thought; I do not need to have any further insight into it' (Kant, *Critique of Pure Reason*, trans. Pluhar, p. 30 [B xxix]). See Hare, 11; Fichte, pp. 245 f. [282–3 § 20 Remark]; Georg Wilhelm Friedrich Hegel, *Elements of the Philosophy of Right*, ed. A. W. Wood and trans. H. B. Nisbet, Cambridge, [1820] 1991, pp. 135–40 [§§ 105–12, with Remarks and Additions]; see *ibid.*, p. 43 [§ 9]; Kant, *Groundwork*, p. 3; Daniel N. Robinson and Rom Harré, 'The Demography of the Kingdom of Ends', *Philosophy*, lxix (1994); J. B. Schneewind, 'Autonomy, Obligation, and Virtue: An Overview of Kant's Moral Philosophy', *Cambridge Companion to Kant*, ed. P. Guyer, Cambridge, 1992; Peter J. Steinberger, 'The Standard View of the Categorical Imperative', *Kant-Studien*, xc (1999), pp. 91, 99; and Bernard Williams, *Ethics and the Limits of Philosophy*, Cambridge, 1985. Also see Henry E. Allison, *Kant's Theory of Freedom*, Cambridge, 1990, pp. 180–98; Karl Ameriks, 'The Hegelian Critique of Kantian Morality', *New Essays on Kant*, ed. B. den Ouden and M. Moen, New York, 1987; Marcia Baron, *Kantian Ethics Almost Without Apology*, Ithaca, 1995; P.-C. Lo, 'A Critical Re-evaluation of the Alleged "Empty Formalism" of Kantian Ethics', *Ethics*, xli (1981); Onora O'Neill, *Acting on Principle*, New York, 1975; Jeffrey Edwards, 'Egoism and Formalism in the Development of Kant's Moral Philosophy', *Kant-Studien*, xci (2000); and Christian J. Onof, 'A Framework for the Derivation and Reconstruction of the Categorical Imperative', *Kant-Studien*, lxxxix (1998).

<sup>12</sup> Often, this is solely attributed to Hegel. See Hegel, *Elements of the Philosophy of Right*, pp. 49 [§ 15 Addition], 113 [§ 81, Remark], 115 f. [§ 82, Addition], 124 f. [§ 99, Remark], 126 f. [§ 100, Remark], and 131 f. [§ 104]. See also *ibid.*, pp. 117 f. [§ 86, Addition] and 104 [§ 73].

<sup>13</sup> See Kant, *Groundwork*, pp. 37–40 [4:428–32]. We should not forget that 'morality is the condition under which *alone* a rational being can be an end in itself, since only through this it is possible to be a lawgiving member in the kingdom of ends' (*ibid.*, p. 42 [4:435]).

universal actions, as it would undermine rights to ownership, liberty, and, most importantly, the *moral* law.<sup>14</sup> As lawbreakers have wrongfully subjected the law-abiding public to harmful actions, Kant believes that it is only just to then subject the former to the harmful maxims they attempted to impose on their community.<sup>15</sup>

This part of Kant's moral theory is seldom disputed. The difficulty is in accurately interpreting the justifications he offers to support the right of the state to punish these criminals for their crimes. The most popular understanding is that Kant's theory of punishment is a retributivist theory, *not* a utilitarian nor a consequentialist theory. Indeed, Kant's political writings contain several criticisms of the use of punishment as a kind of tool to realize future goals.<sup>16</sup> As an example, Kantian punishments must always treat human beings as ends-in-themselves and never as a means to some future goal: '*Punishment by a court (poena forensis)* ... can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society. It must always be inflicted upon him only *because he has committed a crime*'.<sup>17</sup> Punishments are to be administered solely on the basis of criminal guilt, not on consideration of social utility.<sup>18</sup> In addition, the particular form a punishment takes is to be specified by the kind and severity of crime committed. Therefore, a punishment's form should not be influenced by extraneous concerns.

One of the most oft-cited passages of Kant's in favour of the view that he justifies punishment purely on retributive grounds is his 'blood guilt' example in *The Metaphysics of Morals*:

Even if a civil society were to be dissolved by the consent of all its members (e.g., if a people inhabiting an island decided to separate and disperse throughout the world), the last murderer remaining in prison would first have to be executed, so that each has done to him what his deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment;

<sup>14</sup> Kant's moral theory would not tolerate all universal maxims, such as 'destroy human life whenever you please', as the respect for persons must remain primary as a rule. This would contradict Hare's reading. However, Hare is quite correct to point that '[m]oral principles do not have to be as simple and general as Kant seems to have thought, and they can still be universal all the same' (see Hare, 'Could Kant Have Been a Utilitarian?' pp. 7 f.).

<sup>15</sup> See my 'Corlett on Kant, Hegel, and Retribution', p. 563.

<sup>16</sup> 'But to look upon all punishments and rewards as mere machinery in the hands of a higher power, serving only to put rational beings into activity toward their final purpose (happiness) is so patently a mechanism which does away with the freedom of the will that it need not detain us here' (Kant, *Critique of Practical Reason*, ed. and trans. M. J. Gregor, Cambridge, [1788] 1997, p. 35 [5:38] and Kant, *Practical Philosophy*, ed. and trans. M. J. Gregor, Cambridge, 1996, pp. 170 f.). See Kant's essay 'The End of All Things' in Kant, *Religion within the Boundaries of Mere Reason*, ed. A. Wood and G. di Giovanni, Cambridge, 1998, p. 204 [8:338–9].

<sup>17</sup> Kant, *Metaphysics of Morals*, p. 105 [6:331].

<sup>18</sup> Likewise, rule-utilitarians can justify punishments without being retributive.

for otherwise the people can be regarded as collaborators in this public violation of justice.<sup>19</sup>

While all other debts might or might not be forgiven and all other contracts perhaps left unfulfilled, it is only with the punishing of criminals where we are commanded to treat them strictly and severely. The importance Kant gives to punishing violations of justice is no minor concern, for 'if justice goes, there is no longer any value in human beings living on the earth'.<sup>20</sup> In fact, some of our value as persons is partly derived from our sacred defence of justice in human community.

Not to be overlooked, Kant's theory of punishment is often linked with the *lex talionis*: 'whatever undeserved evil you inflict upon another within the people, you inflict upon yourself. If you insult him, you insult yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself'.<sup>21</sup>

Indeed, with regard to executing murderers, Kant in fact argues for 'the strict law of retribution' in '[t]his fitting of punishment to the crime'.<sup>22</sup> This is justified on the grounds that strict justice is not objectionable in the least when the criminal's inner wickedness is particularly noticeable.<sup>23</sup> It must be said that it is rather difficult to reconcile Kant's motto of equity ('the strictest right is the greatest wrong') with this insistence that all murderers are to be killed as punishment for killing someone else.<sup>24</sup> Nevertheless, what is critical for Kant in justifying capital punishment for murderers is the murderers' demonstration of inner wickedness.

<sup>19</sup> Kant, *Metaphysics of Morals*, p. 106 [6:333]. It is easy to forget that we also have a duty to forgive fellow human beings, which at times may result in abandoning the distribution of punishments to culpable violators of moral and positive law (ibid., p. 208 [6:461]).

<sup>20</sup> Ibid., p. 105 [6:332]. Hegel thinks just the opposite: 'Let justice be done' should not have as its consequence 'even if the world should perish' (Hegel, *Elements of the Philosophy of Right*, p. 157 [§ 130]).

<sup>21</sup> Ibid. See Kant, *Lectures on Ethics*, ed. P. Heath and J. B. Schneewind, trans. P. Heath, Cambridge, 1997, pp. 310 f. [§ 49 27:555]. (I am citing a passage from Johann Friedrich Vigilantius's notes from Kant's 1793–4 lecture course entitled 'The Metaphysics of Morals'.) Eugen Dühring argued that the *lex talionis* satisfied too low a penal threshold, as to properly negate an offending will one must cause more harm to the criminal than he or she committed to make up the shortfall of his or her 'lack of regard for others shown by that will in the crime' (Small, 42). At times Kant seems to be somewhat sympathetic to this approach: 'Every deed that violates a human being's right deserves punishment, the function of which is to *avenge* a crime on the one who committed it (not merely to make good the harm that was done)' (Kant, *Metaphysics of Morals*, p. 207 [6:460]).

<sup>22</sup> Kant, *Metaphysics of Morals*, p. 106 [6:332].

<sup>23</sup> Ibid., pp. 106 f. [6:333]. Other interesting examples involve Kant's suggested punishments for rape, pederasty, and bestiality (ibid., p. 130 [6:363]).

<sup>24</sup> Ibid., p. 27 [6:235].

Perhaps because of its discord with modern notions of morality, one aspect of Kantian punishment that is little mentioned in the literature is the fact that, for Kant, punishment is a physical harm, properly understood.<sup>25</sup> Psychological punishments or non-painful punishments are not seriously considered for most crimes.<sup>26</sup> In fact, the physical infliction of punishment upon a criminal is a necessary consequence in the allocation of justice, even if the criminal transgression precipitating the punishment is not naturally connected with moral wickedness.<sup>27</sup> In a sense, the criminal ‘pays’ for the crime she committed by experiencing a certain amount of, potentially lethal, pain.<sup>28</sup> This is not to imply that Kant was some kind of sadist – to be clear, he was certainly not.<sup>29</sup> In fact, he finds public rewards offered by the government to be ‘more in harmony with morality’ than punishments.<sup>30</sup>

I do not believe that this evidence is enough to justify labelling Kant a retributivist, for his theory seems to also go in another direction: He is quite explicit in stating that governments can and do only impose deterrent punishments.<sup>31</sup> For Kant, governments are concerned primarily with pragmatic matters, rather than with matters of justice, all things considered. One example that has been ignored by nearly all commentators is the fact that in certain instances no punishment should be inflicted, even where the guilty parties *deserve* to be punished.<sup>32</sup> In the ‘Doctrine of Right’, Kant supposes that we have before us murderers and their accomplices. He tells us that the idea of judicial legal authority dictates that all of these persons must be executed, in accordance with universal laws – the categorical imperative regarding punishment. Nevertheless, Kant asserts that

<sup>25</sup> Kant, *Critique of Practical Reason*, p. 34 [5:37] and Kant, *Lectures on Ethics*, trans. Heath, pp. 79 [27:286], 304 [§ 38 27:547], 308 f. [§§ 43–4 27:552 f.], 312 [§ 50 27:556]. (From the *Lectures*, the first citation is from the section ‘Of Rewards and Punishments’ from Georg Ludwig Collins’s 1784 notes. The following citations are Vigilantius’s notes.)

<sup>26</sup> There is an exception for crimes where the appropriate manner of punishing criminals is to levy fines. Moreover, wealthy criminals who will not be ‘hurt’ by the fine are to be publicly embarrassed, perhaps a form of psychological punishment. (Kant, *Metaphysics of Morals*, p. 106 [6:332].)

<sup>27</sup> Kant, *Critique of Practical Reason*, pp. 34 f. [5:37] and Kant, *Metaphysics of Morals*, p. 104 [6:331].

<sup>28</sup> See Tom Sorell, *Moral Theory and Capital Punishment*, Oxford, 1987, p. 134.

<sup>29</sup> ‘No punishment should be coupled with cruelty’ Kant, *Lectures on Ethics*, trans. Heath, p. 311 [§ 49 27:556].

<sup>30</sup> Kant, *Lectures on Ethics*, trans. L. Infield, Indianapolis, 1963, p. 57.

<sup>31</sup> Ibid., p. 55 and Kant, *Lectures on Ethics*, trans. Heath, p. 79 [27:286] in the section ‘Of Rewards and Punishments’. See the section entitled ‘Of the Lawgiver’ in *ibid.*, p. 79 [27:286]. (Both are from the notes of G. L. Collins.) This statement mirrors Fichte’s belief that the purpose of punishment is solely to deter. (See Fichte, pp. 226–48 [260–85 § 20].)

<sup>32</sup> I know of only three treatments: Byrd, 196, 196n144; Scheid, 268–71; and Tunick, 63. Corlett takes this view only with regard to his own reinterpretation of a Kantian penal theory (Corlett, 78, 78n7).

should the number of criminals be so great a number that the state would be left without any subjects if they were all executed, 'the sovereign must also have it in his power, in this case of necessity (*casus necessitatis*), to assume the role of judge (to represent him) and pronounce a judgement that decrees for the criminals a sentence other than capital punishment, such as deportation, which still preserves the population'.<sup>33</sup>

In this example, the preservation of the community outweighs retributive concerns when administering particular punishments – in violation of much of what form punishment ought to take as a categorical imperative.

Kant's use of the sovereign offers another example that does not support a retributivist reading. First, he declares that 'the presently existing legislative authority ought to be obeyed, whatever its origin', even though this authority may be completely at odds with reason or morality.<sup>34</sup> People would then be forced to live according to laws that may violate universality, for instance. In addition, citizens have no rights to sedition or rebellion – even if the sovereign abuses his authority. Any attempt to do either is high treason, punishable by state execution.<sup>35</sup> Kant says: '[l]ike a chasm that irretrievably swallows everything, the execution of a monarch seems to be a crime from which *the people* cannot be absolved, for it is as if the state commits suicide'.<sup>36</sup> Finally, not only is the sovereign immune from legal punishment as king, but even if he should be dethroned, he cannot be punished for any actions he performed while a sovereign.<sup>37</sup>

Even when reiterating his commitment to the *lex talionis*, Kant tells us that punishments ought to be less severe than the criminal deserves when the punishment would be a crime against humanity: Those convicted of rape or pederasty should be castrated, those of bestiality expelled from the community for life, etc.<sup>38</sup> His argument is that we may forgo punishing to the letter of the law, so long as the penalty we choose is commensurable to its spirit.<sup>39</sup>

<sup>33</sup> Kant, *Metaphysics of Morals*, p. 107 [6:334].

<sup>34</sup> Ibid., p. 95 [6:319]. Should a citizen resist the rule of his sovereign, she 'would be punished, got rid of, or expelled in accordance with the laws of this authority, that is, with every right' (ibid.).

<sup>35</sup> Ibid., p. 96 [6:320]. He says further: 'It is the formal *execution* of a monarch that strikes horror in a soul filled with the idea of human beings' rights, a horror that one feels repeatedly as soon as and as often as one thinks of such scenes as the fate of Charles I or Louis XVI' (ibid., p. 97n [6:320]).

<sup>36</sup> Ibid., pp. 97n–98n [6:320] (emphasis added). As 'the people' are implicated in a crime most, if any at all, did not participate in in any way, Kant sanctions yet again the violation of a universal maxim.

<sup>37</sup> Ibid., pp. 95, 98 [6:319, 323].

<sup>38</sup> Ibid., p. 130 [6:363].

<sup>39</sup> Ibid.



The most well-known example of Kant's which appears to justify the primacy of consequentialist above retributive concerns regards two shipwrecked men fighting over a raft that will only hold one of them.<sup>40</sup> Should one of them push the other off the raft – drowning him – in order to save his own life, the surviving man is 'to be morally condemned but not legally punished'.<sup>41</sup> Kant's reason is explicit:

the punishment threatened by the law could not be greater than the loss of his own life. A penal law of this sort could not have the effect intended, since a threat of an ill that is still *uncertain* (death by a judicial verdict) cannot outweigh the fear of an ill that is *certain* (drowning). Hence the deed of saving one's life by violence is not to be judged *inculpable* (*inculpabile*) but only *unpunishable* (*impunibile*).<sup>42</sup>

The surviving person is to be relieved of legal punishment in part due to his fulfilling a duty to preserve his own life.<sup>43</sup> In addition, the surviving person avoids punishment only because punishing him would not have the desired effect that punishments ought to have, derived from 'threat'. Kant argues also against punishing mothers who kill their illegitimate children and soldiers who murder each other in duels on the same grounds: These persons ought not to be punished for doing so would not *deter* others in committing the same actions in the future.<sup>44</sup>

At first glance, it might appear as if Kant lacks a coherent theory of punishment. As is well known, in some instances Kant seems to be concerned solely with retributive justifications for punishing and for particular punishments. For example, recall the illustration of the island community about to dissolve with a murderer left in its prison: The community *must* execute the murderer for the sake of justice or otherwise the community shares in the murderer's guilt. Almost to the contrary, I have shown that there are many other instances where Kant seems to justify punishing criminals solely upon consequentialist grounds, where if persons cannot be deterred by the threat of punishment, they ought not to be punished at all. I would propose that the apparent contradiction is due mostly to a misreading of Kant, for when

<sup>40</sup> See Kant, *Metaphysics of Morals*, p. 28 [6:235 f.].

<sup>41</sup> I take this phrase from Tunick, 65.

<sup>42</sup> Kant, *Metaphysics of Morals*, p. 28 [6:235 f.] (emphasis given). See Jeffrie G. Murphy, *Kant: The Philosophy of Right*, London, 1970, p. 130n3.

<sup>43</sup> Kant, *Groundwork*, p. 11 [4:397]. Fleischacker argues that in this example 'Kant does not hold that the impossibility of deterring someone morally justifies not punishing him; on the contrary, he stresses the fact that deterrence concerns only the subjective (empirical), not the objective value of punishment' (Fleischacker, 194). The problem with this view is that Fleischacker has misrepresented Kant's penology by not taking into consideration that punishment is itself an empirical practice – and, for Kant, governments can only administer *deterrent* punishments.

<sup>44</sup> Kant, *Metaphysics of Morals*, pp. 108 f. [6:336 f.].

Kant is writing in favour of one version of punishment rather than another he is doing so in very different contexts.

### III. TWO SPHERES OF LAW

For Kant, there is a dichotomy of juridical (i.e. positive) law and moral law. Juridical laws concern themselves solely with external actions and their conformity to a government's legislation – which may or may not be rational.<sup>45</sup> In order to follow a juridical law we must act solely in accordance with written legislation governing a human community. In so doing, our actions may be perfectly legal but not always commensurable with standards set by morality. Juridical laws require incentives other than moral duty, as this form of duty is an internal incentive.<sup>46</sup> As an example, Kant says: 'It is an external duty to keep a promise made in a contract; but the command to do this merely because it is a duty, without regard for any other incentive, belongs to *internal* lawgiving alone'.<sup>47</sup>

When we act from duty alone, we act with respect to morality and the moral law. Acting otherwise is primarily a matter for juridical law.<sup>48</sup>

As we have seen previously, for Kant, the moral law is first and foremost expressed in the categorical imperative.<sup>49</sup> This imperative commands us to act only in such a way that our maxim – expressed by our action – can be a universal law.<sup>50</sup> As rational beings, we act within a 'kingdom of ends' whereby we subject ourselves to universal laws at one and the same time as we perform particular actions.<sup>51</sup> Most importantly, this law becomes the standard by which we may judge the moral correctness of every particular action. Before we perform any action we must first consider – consciously or unconsciously – the

<sup>45</sup> See Kant, *Metaphysics of Morals*, pp. 14, 17 [6:220, 224]. As positive laws may be irrational, Kant's argument against the right to sedition or rebellion against the state may be problematic (see *ibid.*, pp. 96 f. [6:320]). In addition, human beings are only ends in themselves in morality, not in positive law (Kant, *Groundwork*, p. 42 [4:435]). This is not to say that Kant's project of a kingdom of ends is complete fantasy, for one of the primary foundations of a kingdom of ends is its possible realization through human conduct (*ibid.*, p. 44n [4:437]). Kant hints that there would be no need for punishment in the ideal society (Kant, *Critique of Pure Reason*, trans. N. K. Smith, London, 1963, A317/B373; see Fleischacker, 191 and Fichte, 244 [§ 20 282V(f.)]).

<sup>46</sup> Kant, *Metaphysics of Morals*, p. 20 [6:219]. See *ibid.*, p. 23 [6:230].

<sup>47</sup> *Ibid.*, p. 21 [6:220].

<sup>48</sup> See *ibid.*, p. 20 [6:219].

<sup>49</sup> *Ibid.*, p. 15 [6:222–3]. See *ibid.*, p. 19 [6:227] and Kant, *Groundwork*, pp. 5 [4:392]; 30–3 [4:420–4] and 39 f. [4:430–2].

<sup>50</sup> Kant, *Groundwork*, pp. 15 [4:402], 16 [4:403], and 31 [4:421].

<sup>51</sup> *Ibid.*, pp. 41 [4:432] and 42 [4:434]. See *ibid.*, p. 40 [4:432].

effect on society should all its members adopt the maxim on which we are intending to act. For example, if we were to consider lying, stealing, or committing a murder, we are to reject this behaviour on the grounds that everyone might lie, steal, or murder each other, bringing about a world without trust or promises, protection of property, or basic human liberty. By acting in accordance with morality, a human being can fully exist as an end-in-itself, 'since only through this is it possible to be a lawgiving member in the kingdom of ends'.<sup>52</sup>

Were we to punish criminals from this purely moral perspective, it would be essential that we justify punishment primarily as an act of justice: 'this constitutes what is essential in this concept [of punishment]'.<sup>53</sup> In other words, punishment must be primarily commensurable with justice in order for it to be distributed to a particular person. Therefore, all punishments ought to be given only to the guilty and within certain limits. With regard to the violation of moral laws, Kant is clearly in favour of retributive measures of punishing criminals. On the other hand, from the standpoint of juridical, or positive, law, Kant is concerned primarily with deterrence-based justifications for punishment. This crucial dichotomy is rarely appreciated by commentators on Kant's political philosophy.<sup>54</sup>

It is important to note that this matter – which he called a 'quandary' – was not left unnoticed by Kant, for he says:

The knot can be undone in the following way: the categorical imperative of penal justice remains (unlawful killing of another must be punished by death); but the legislation itself (and consequently also the civil constitution), as long as it remains barbarous and undeveloped, is responsible for the discrepancy between the incentives of honour in the people (subjectively) and the measures that are (objectively) suitable for its purpose. So the public justice arising from the state becomes an *injustice* from the perspective of the justice arising from the people.<sup>55</sup>

<sup>52</sup> Ibid., p. 42 [4:435]. See *ibid.*, p. 15 [4:403] and Kant, *Metaphysics of Morals*, p. 182 [6:429].

<sup>53</sup> Kant, *Critique of Practical Reason*, p. 34 [5:37].

<sup>54</sup> See Kant, *Metaphysics of Morals*, pp. 17, 20 [6:219, 225] and 10, *noted*. The best treatment of Kant's theory of punishment is Tunick, 60–78, esp. 77 f. A fairly similar dichotomy is present in Hegel (see Hegel, *Elements of the Philosophy of Right*, p. 121 [§ 94 Addition]). Also see Hare, 11 and Kant, *Groundwork*, pp. 57 [4:452] and 61 [4:457].

<sup>55</sup> Kant, *Metaphysics of Morals*, p. 109 [6:336–7]. See Allen W. Wood, *Kant's Ethical Thought*, Cambridge, 1999, p. 322: 'Kant holds that we have a *moral* obligation to limit ourselves to actions that are right [R], but that duty is no part of R itself. R grounds only *juridical* duties, which are distinguished from ethical duties by the fact that their concept contains no determinate incentive for complying with them'. Unfortunately, there is no discussion of the role(s) of moral and juridical duties regarding punishment. (*Ibid.*, p. 407n32 and Allen W. Wood, 'Kant's Practical Philosophy', *Cambridge Companion to German Idealism*, ed. Karl Ameriks, Cambridge, 2000.)

From the standpoint of the moral law, the positive law may well be inadequate. Indeed, we might best evaluate juridical law via moral law. However, this is just one standpoint. The other standpoint is that of the state which deals with practical matters of external actions, which it must try to influence. In practical concerns, such as state legislation, justice is served only when the letter of the law is observed. For Kant, the best of all possible human communities will be one that fully develops itself when it perfectly unifies the spheres of juridical and moral law.

This noble ideal was not a reality in Kant's time nor is it true today. Human societies will remain in a 'barbarous and undeveloped' condition until they are able to commensurate juridical and moral law. Thus, our attempts to determine the most just punishment for a particular crime begin in an undeveloped condition, where the consequential effects of a punishment take priority over the satisfaction of purely retributive demands. We may improve our attempts over time by greater incorporation of retributive standards into our theory of punishment. Crucially, while the form a punishment takes may eventually begin to prioritize retribution over consequences, punishment takes a consequentialist form at first, as it attempts to shape external human behaviour. Punishing criminal conduct strictly in accordance with what the moral law dictates is the ultimate aim of the practice, but not where we begin.<sup>56</sup>

Kant states in his 1792 letter to Johann Benjamin Erhard:

In a world of moral principles governed by God, punishments would be categorically necessary (insofar as transgressions occur). But in a world of moral principles governed by men, the necessity of punishments is only hypothetical, and that direct union of the concept of transgression with the idea of deserving punishment serves the ruler only as a prescription for what to do. So you are right in saying that the *poena meremoralis* ['ethical penalty'] (which perhaps came to be called *vindicativa* ['avenging punishment'] for the reason that it preserves the divine justice), even if its goal is merely medicinal for the criminal and the setting of an example for others, is indeed a *symbol* of something deserving punishment, as far as the condition of its authorization is concerned.<sup>57</sup>

In this passage, Kant agrees with Ernst Ferdinand Klein that 'punishment is the symbol of an action's deserving punishment, by

<sup>56</sup> For example, Kant tells us that the legal system he is describing is that of 'the state in idea, as it ought to be in accordance with pure principles of right' (Kant, *Metaphysics of Morals*, p. 90 [§ 45 6:313]).

<sup>57</sup> Kant, *Kant: Philosophical Correspondence, 1756–99*, ed. and trans. A. Zweig, Chicago, 1967, p. 199 [11:398–9] and Kant, *Correspondence*, ed. and trans. A. Zweig, Cambridge, 1999, p. 448. See Kant, *Metaphysics of Morals*, pp. 24–7, 104–6 [6:231–4, 331 f.]. Fleischacker cites only the first half of this letter, which may have further contributed to his misinterpretation of it (Fleischacker, 204).

means of a mortification of the criminal that corresponds to the crime committed'.<sup>58</sup> More importantly, Kant agrees with the interpretation of his penal theory being offered, as he admits that 'the necessity of punishments is only hypothetical' with regard to 'a world of moral principles governed by men'. Rather than justifying punishments completely devoid of retributive concerns, Kant argues instead that these concerns ought to be considered as 'a prescription for what to do'. Thus, Kant *tempers* a consequentialist theory of punishment by moulding punishment in a shape that would fulfil retributive standards.<sup>59</sup> However, this matter is one of retributive concerns helping to form what is primarily consequentialist punishment and not *vice versa*.

Furthermore, I would like to suggest that when Kant speaks of sovereigns<sup>60</sup> 'performing injustice in the highest degree' when offering pardons or reduced sentences to criminals, the injustice in question is moral, not juridical.<sup>61</sup> Such a matter is related solely to the moral law existing internally in human beings, which we can discern barely and with much difficulty. From the standpoint of juridical law the sovereign who offers pardons acts justly if this pardon positively colours certain consequences. In fact, it would be a grave injustice if a sovereign did not pardon criminals where carrying out their sentence would lead to the breakdown of the community.<sup>62</sup>

<sup>58</sup> See Kant, *Kant: Philosophical Correspondence*, p. 199n3. Kant says of Klein on criminal right: 'Most of what he says is excellent and quite in accord with my own view' (ibid., p. 199).

<sup>59</sup> As he says in the letter to Erhard, the authorization of punishment is dependent upon retributive justification (esp. its concept of desert), subsumed under a primarily consequentialist theory of punishment.

<sup>60</sup> Kant's use of 'sovereign' is a term easy to confuse. In the *Groundwork*, the 'sovereign' is each rational being in the kingdom of ends: 'He belongs to it *as sovereign* when, as lawgiving, he is not subject to the will of any other' (Kant, *Groundwork*, p. 41 [4:433]). However, in *Metaphysics of Morals*, the sovereign is a single person who alone rules over his community (Kant, *Metaphysics of Morals*, pp. 95–8 [6:319–23]). I believe that 'sovereign' in the first sense is related to individual autonomy and in the second it is related to a particular individual (i.e. a monarch).

<sup>61</sup> Kant, *Metaphysics of Morals*, p. 109 [6:337]. See Don Becker, 'Kant's Moral and Political Philosophy', *Routledge History of Philosophy*, vol. 6, *The Age of German Idealism*, ed. R. C. Solomon and K. M. Higgins, London, 1993, p. 84. I believe that Jeffrie Murphy was mistaken in declaring that Kant held no such view, as Kant says: 'conformity with juridical laws is the *legality* of an action and conformity with ethical laws is its *morality*' (Kant, *Metaphysics of Morals*, pp. 14 [6:220] and 17 [6:225]; see Jeffrie G. Murphy, 'Kant's Concept of A [sic] Right Action', *Kant Studies Today*, ed. Lewis W. Beck, La Salle, 1969, pp. 471–95).

<sup>62</sup> Kant, *Metaphysics of Morals*, p. 107 [6:334]. To recall Kant's 'blood guilt' example, it would then seem to follow that the last murderer in prison might not be executed upon the dissolution of his island community if the community's preservation is no longer at issue: When its preservation may collapse, alternative penal remedies are justified to continue the community.

## IV. THE PROBLEM OF INTENTIONALITY

While they are subject to the particular laws of their place of residence, Kant believes that human beings ought to follow the moral law at all times.<sup>63</sup> The moral law is not some utopian ideal we might aspire to, yet never fulfil. Instead, the significance given to the moral law via reason is that 'the moral law is solely practical', its realization by human beings is a true possibility and not a product of wishful thinking.<sup>64</sup> It is the moral law which determines the concept of right and wrong, as a standard by which we fine-tune legislation to unite as closely as possible juridical law with the moral law.<sup>65</sup>

Rather confusingly, Kant suggests that when we violate the moral law our transgression deserves punishment.<sup>66</sup> In this way, Kant speaks of 'punishment (*poena*)' resulting from the moral law (the categorical imperative)'s violation as 'the rightful effect'.<sup>67</sup> Specifically, the 'rightful effect' is 'what is culpable is punishment'.<sup>68</sup> This matter might be understood in one of two ways. One interpretation is that persons who violate the moral law ought to be punished for their violation – punishment is the 'rightful effect' – but these persons may avoid being punished because they have not violated the juridical law. The difficulty with this view is that it makes Kant's proposition that criminals should be punished on the basis of *mens rea* appear contradictory.<sup>69</sup>

<sup>63</sup> See Kant, *Critique of Practical Reason*, pp. 35 f. [5:38], 70 f. [5:82–4] and 79 [5:93]; Kant, *Metaphysics of Morals*, p. 9 [6:215].

<sup>64</sup> Kant, *Critique of Practical Reason*, p. 44 [5:50]. See *ibid.*, pp. 33, 74 [5:37, 86–7]; Kant, *Metaphysics of Morals*, pp. 42 f. [6:252–3]; Kant, 'On the Proverb: That May be True in Theory, But Is of No Practical Use', *Perpetual Peace and Other Essays on Politics, History, and Moral Practice*, ed. T. Humphrey, Indianapolis, 1983, p. 71.

<sup>65</sup> See Kant, *Critique of Practical Reason*, p. 54 [5:63]; Kant, *Lectures on Ethics*, trans. Heath, pp. 73 f. [27:279] ('*De Littera Legis*'), 80 [27:287] ('Of Rewards and Punishments'), 83 [27:291] ('Of Degrees of Imputation'); and Kant, *Metaphysics of Morals*, p. 16 [6:223 f.]. (All citations from Kant's *Lectures* are from the notes of G. L. Collins.) See also Kant's discussion of public right at Kant, *Metaphysics of Morals*, p. 89 [6:311]; 'On a Supposed Right to Lie because of Philanthropic Concerns', p. 165 [429]. In Kant, *Ethical Philosophy*, 2nd edn., ed. and trans. J. W. Ellington, Indianapolis, 1994: in the move 'from a metaphysics of right (which abstracts from all empirical determinations) to a principle of politics (which applies these [metaphysical] concepts [of right] to instances provided by experience' we 'gain the solution of a problem of politics in accordance with the universal principle of right'. (Brackets supplied by Ellington.)

<sup>66</sup> Kant, *Critique of Practical Reason*, p. 34 [5:37]. In 'Of Rewards and Punishments', Kant says: 'Punishment in general is the physical evil visited upon a person for moral evil'. (See Kant, *Lectures on Ethics*, trans. Heath, p. 79 [27:286]. See pp. 308 [§ 43 27:552] ('In punishments, a physical evil is coupled to moral badness'), 309 [§ 44 27:553].)

<sup>67</sup> Kant, *Metaphysics of Morals*, p. 19 [6:227] (emphasis given). See Kant, *Lectures on Ethics*, trans. Heath, pp. 76 [27:282], 80 f. [27:288], and 306 [§ 40 27:549].

<sup>68</sup> *Ibid.*

<sup>69</sup> As an example, see *ibid.*, pp. 106 f. [6:333].

On the other hand, I believe that this problem is best understood as an argument for the importance of intentionality for Kant in the attribution of punishments. This is a new and crucial problem in coming to grips with Kant's theory of punishment. Kant insists: 'the state of mind of the subject, whether he committed the deed in a state of agitation or with cool deliberation, makes a difference in imputation, which has results'.<sup>70</sup> This state of mind – the person's will – is what 'always takes first place in estimating the total worth of our actions and constitutes the condition of all the rest'.<sup>71</sup> Whatever action we perform is granted some level of moral worth from the commensurability this action has with the maxim employed when justifying the particular action. The purpose, or consequence, pursued is irrelevant for consideration.<sup>72</sup>

When we legally punish, we ought to appraise not only the crime committed, but the 'inner wickedness' of the criminal when performing an offence.<sup>73</sup> While juridical laws bear on the external actions of persons independent of the reasons for acting – as moral laws are concerned with the state of mind and decision-making process of human beings – a person's intentional (i.e. culpable) disposition when performing actions which violate the positive law is an important factor in attributing *punishable* guilt.<sup>74</sup> Individuals who deliberately breach these laws are more culpable for their criminal action(s) than those who do so unintentionally.<sup>75</sup>

Of course, we should not always assume that a criminal's intentionality is united in his criminal action(s), for only in certain instances are the two fully connected.<sup>76</sup> On some occasions, actions may become criminal due to unintended consequences. This gap between the perceived (our physical movements) and the unperceived (the state of our moral disposition) marks the difficulty of this project of calibrating punishments to persons solely regarding their moral guilt.<sup>77</sup> It is precisely this gap between the two which needs to be better

<sup>70</sup> Kant, *Metaphysics of Morals*, p. 20 [6:228].

<sup>71</sup> Kant, *Groundwork*, p. 10 [4:397].

<sup>72</sup> *Ibid.*, p. 13 [4:399 f.]. See *ibid.*, p. 14 [4:401].

<sup>73</sup> Kant, *Metaphysics of Morals*, pp. 106 f. [6:333].

<sup>74</sup> See Kant, *Lectures on Ethics*, trans. Heath, p. 74 [27:280] in the section '*De Littera Legis*'.

<sup>75</sup> I am referring to the common legal distinctions of *actus reus* and *mens rea* and their often critical importance in attributing punishments to criminals.

<sup>76</sup> Kant, *Metaphysics of Morals*, p. 108 [6:335]. See Kant, *Groundwork*, pp. 13 [4:399 f.] and 19 f. [4:407].

<sup>77</sup> It should also be kept in mind that the calibration of juridical law towards encompassing moral law is not a project to be completed overnight. This process may well take generations. (See Kant, 'Idea for a Universal History with a Cosmopolitan Intent', Kant, *Perpetual Peace and Other Essays*, Akademie p. 19.)

understood if we are to impart criminal guilt and/or penalties to criminals due, at least in part, to their intentional disposition.<sup>78</sup>

In the *Critique of Pure Reason*, Kant says:

The real morality of actions, their merit or guilt, even that of our own conduct, thus remains entirely hidden from us. Our imputations can refer only to the empirical character. How much of this character is ascribable to the pure effect of freedom, how much to mere nature, that is, to faults of temperament for which there is no responsibility, or to its happy constitution (*merito fortunae*), can never be determined, and upon it therefore no perfectly just judgements can be passed.<sup>79</sup>

He states elsewhere: 'In fact, it is *absolutely impossible* by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty'.<sup>80</sup>

Morality is poised at the very boundary of human knowledge: 'reason would overstep all its bounds if it took upon itself to explain how reason can be practical'.<sup>81</sup> What Kant fails to sufficiently consider is the fact that the moral law's true content – which we have no knowledge of – might actually vary greatly with his theory of moral law.

While Kant denies that we could ever know for sure whether or not a person was morally motivated, the moral law can only have authority over our wills when it is possible for us to be motivated by it:<sup>82</sup> for 'what counts is not actions, which one sees, but those inner principles of actions that one does not see'.<sup>83</sup> Simply put, the essential

<sup>78</sup> As Kant says, 'Right [justice] must never be adapted to politics; rather, politics must always be adapted to right' (Kant, *Ethical Philosophy*, p. 166 [429]). See Kant, *Lectures on Ethics*, trans. Heath, p. 306 [§40 27:550]. The fact that most interpretations of Kant's theory of punishment have not found such a gap previously ought then to be seen as one instance in which many were misled.

<sup>79</sup> Kant, *Critique of Pure Reason*, trans. N. K. Smith, p. 475n [A552/B580] (emphasis added). See Sullivan, *Immanuel Kant's Moral Theory*, p. 243.

<sup>80</sup> Kant, *Groundwork*, p. 19 [4:407]. Due to our inability of knowing dispositions with any certainty, T. H. Green criticizes Kant for continuing to include intentionality in his theory of punishment. Green does what he believes Kant ought to have done: base his theory of punishment on more consequentialist grounds. (See Thomas Hill Green, *Lectures on the Principles of Political Obligation*, London, [1882] 1941, pp. 180–205 [sections 176–206] and my 'T. H. Green's Theory of Punishment', *History of Political Thought* (forthcoming).)

<sup>81</sup> Kant, *Groundwork*, p. 62 [4:458–9]. Kant says: 'We see this as soon as we become convinced that there is a use of pure reason which is practical and absolutely necessary (viz., its moral use). When used practically, pure reason inevitably expands and reaches beyond the bounds of sensibility' (Kant, *Critique of Pure Reason*, trans. Pluhar, p. 27 [B xxv]; see Katrin Flikschuh, *Kant and Modern Political Philosophy*, Cambridge, 2000, p. 194).

<sup>82</sup> I paraphrase Korsgaard, 'Introduction', p. xxi.

<sup>83</sup> Kant, *Groundwork*, p. 20 [4:407]. Nevertheless, this state of affairs does not stop Kant from stating that: 'the pure thought of duty and in general of the moral law, mixed



good in an action ‘consists in the disposition’.<sup>84</sup> Accordingly, judges will have little recourse other than to ‘construe the criminal’s act as intend[ing] to express a universal law even though it probably in fact results from an irrational maxim’.<sup>85</sup> If we are to attribute penal sanctions on the basis of criminal intent – never sure if we are correct – then it may well be that Kant’s theory of punishment would justify the legal punishment (and possible execution) of innocent persons.<sup>86</sup>

Kant admits, ‘only the moral relations of *human beings to human beings* are comprehensible to us’ as an empirical phenomenon.<sup>87</sup> This statement underlies the dilemma at hand. At the very least, neither Kant nor many of ourselves would be satisfied with a system of juridical law which lacked coherence and broad notions of justice, in particular. Yet, the shape which coherence and notions of justice ought to take are quite difficult to ascertain. My reading understands Kant to agree with this view and further argue that we are able to work out ideal notions of justice via theoretical tools, such as the categorical imperative. The problem is how best to incorporate this ideality with the empirical world in which we live. It is beyond the scope of this article to identify an acceptable account of intentionality; suffice to say that while Kant does not rule out the impossibility of such a project, he is aware of its difficulty. This has not lost its strength since Kant’s

with no foreign addition of empirical inducements, has by way of reason alone an influence on the human heart so much more powerful than all other incentives, which may be summoned from the empirical field’ (ibid., pp. 22 f. [4:410–11]).

<sup>84</sup> Ibid., p. 27 [4:416].

<sup>85</sup> See Fleischacker, 202. Fleischacker believes that the Kantian judge ‘appears to be a rather strange human being, and indeed it is not clear that he is, in his formal role, a human being at all’ (ibid., p. 203). Also see where Kant says: ‘Rewards and punishments are merely subjective motivating grounds; if objective grounds no longer avail, the subjective serve merely to replace the want of morality’ (Kant, *Lectures on Ethics*, trans. Heath, p. 80 [27:287]; see p. 284 [§ 24 27:522]; Manfred Kuehn, *Kant: A Biography*, Cambridge, 2001, p. 41; and George, p. 149n39).

<sup>86</sup> This is not to say that Kant would find it desirable to punish innocent persons, for, in fact, the moral law prohibits this from happening. I would argue that the attempt to administer punishments only to guilty persons is one example of how consequentialist punishments are to be tempered by dictates of rational morality to the best degree possible (and for good reason). However, the inability to always (or mostly) positively identify one’s moral disposition when breaching juridical laws plausibly opens the door wide to mistaking unintended consequences for criminal culpability. Thus, the Kantian believes the innocent man being punished is innocent, whereas the ‘naïve utilitarian’ knows the person to be punished is innocent and punishes him anyway. (See my ‘Gilligan on Deterrence and the Death Penalty: Has Legal Punishment Failed Us?’, *Ethics and Justice*, iii/iv (2001/2002); my ‘Utilitarianism, Capital Punishment, and Innocent Persons: A Defense of Bentham’, *Review Journal of Philosophy and Social Science*, special issue, xxvii (2002); and Fred Rosen, ‘Utilitarianism and the Punishment of the Innocent: The Origins of a False Doctrine’, *Utilitas*, ix (1997). On ‘naïve utilitarianism’, see my ‘Corlett on Kant, Hegel, and Retribution’, pp. 578–80.)

<sup>87</sup> Kant, *Metaphysics of Morals*, p. 232 [6:491].

time, as it would appear that just about every project seeking to unify moral law and positive law faces just this dilemma.

## V. CONCLUSION

Interpretations of Kant's theory of punishment have been hampered by the inability of commentators to distinguish properly between Kant's distinction between juridical and moral law. Kant is not retributivist at a primary level and then open to considerations of social utility at a secondary level, affixing specific punishments to particular criminals. Instead, when Kant is arguing on behalf of retributive or deterrent-based justifications he does so from different standpoints. The retributive Kant is concerned with transgressions of the moral law, the consequentialist Kant with juridical law. Indeed, it is easy to confuse the two as Kant does seem to take into account moral culpability when determining the severity of punishment for criminals.

Some commentators, such as Kevin Thompson, are wary of any conflation of moral and positive law: 'the very spheres Kant wished to separate'.<sup>88</sup> On the contrary, I believe that Kant is 'searching for a way to unite the moral law with empirical institutions of justice'.<sup>89</sup> Kant's penal consequentialism is *tempered* by its aspiring desire to conform to the moral law, a project severely curtailed by our inability to know with certainty the actual content of morality aside from a reasonable estimation.<sup>90</sup> This is particularly troublesome as we are supposed to impart punishments to criminals commensurable with their criminal culpability. To compound this problem, by advocating in advance the moral measure to which potential criminals will be held to prior to actual events, Kant's theory of punishment may well impugn penalties to undeserving persons.

Rather than this being a weakness of Kant's theory of punishment, Kant demonstrates instead the weakness of all such theories where criminal culpability is central to the determination of justice. Our inability of knowing with certainty a person's disposition when committing an act is a reality, but not a hindrance. We decide criminal guilt according to standards such as 'beyond a reasonable doubt'. The link between intentions, actions, and criminal culpability remains in

<sup>88</sup> Kevin Thompson, 'Kant's Transcendental Deduction of Political Authority', *Kant-Studien*, xcii (2001), 78.

<sup>89</sup> Lindstedt, 133.

<sup>90</sup> I am therefore at odds with Hare's assertion that Kant 'was a sort of utilitarian, namely a rational-will utilitarian', although I am generally sympathetic with Hare's analysis (Hare, 4).

modern jurisprudence despite our lack of certainty in identifying dispositions with absolute certainty. Thus, Kant's problem remains a contemporary dilemma and not an issue of purely historical importance.<sup>91</sup>

tbrooks@web.de

<sup>91</sup> Presented at the Fourth European Congress for Analytic Philosophy at Lund University, the Scottish Postgraduate Philosophy Association at Stirling University, and the senior postgraduate seminar at the University of Sheffield. I am most grateful to Gustav Arrhenius, Albert Atkin, Jes Bjarup, Meagan Brooks, Andy Clark, Martin Golding, Gerry Hough, David Liggins, Brian O'Connor, and, especially, Fabian Freyenhagen, Bob Stern, and Leif Wenar for very helpful comments on earlier drafts.