

Kant's Justification of the Death Penalty Reconsidered

BENJAMIN S. YOST

Providence College, Providence, RI

Introduction

It is hard to know what to think about Kant's 'passionate sermons' on capital punishment.¹ Kant clearly feels that it is one of the most important punishments in the state's arsenal. But his vehement insistence on the necessity of execution strikes many readers as philosophically suspect. Critics argue that Kant's embrace of the death penalty is incompatible with, or at least not required by, the fundamental tenets of his moral and legal philosophy (Schwarzschild 1985; Merle 2000; Potter 2002; Hill 2003). These arguments typically employ one of two strategies. The first is to deny that execution is required by retribution in even a *prima facie* sense; arguments along this line typically question the coherency of Kant's doctrine of the *ius talionis* (Sarver 1997). The second is to show that there are inviolable moral principles that render the death penalty illegitimate; this criticism typically appeals to the value of human dignity or the right to life (Radin 1980; Pugsley 1981; Schwarzschild 1985; Merle 2000; Potter 2002). There is a third strategy that could be used to criticize Kant, although it is not aimed at him specifically. This strategy invokes legal principles of fairness or due process. It asserts that, regardless of capital punishment's moral appropriateness or legitimacy, capital punishment is unjust because of the fallibility of legal actors and institutions (Nathanson 2001). Someone adopting the third strategy might claim that, while Kant's justification may be acceptable in principle, it fails to justify the death penalty in the world we live in.

It has been said that arguments against Kant's justification of capital punishment are so successful that 'today there is barely an

interpreter who will take a stand in favor of this part of Kant's theory of penal law' (Merle 2000: 326–7). In this paper, I will take such a stand, arguing that Kant does have a viable justification of capital punishment, one that coheres with the basic tenets of his practical philosophy and generates a pressure to carry out executions.

Since Kant's work is the locus classicus of retributivist justifications of capital punishment, this paper has implications for broader debates in the philosophy of punishment. However, my primary goal is exegetical. Kant's justification is not obvious; its main points are not explicitly defended, and some of what Kant does say about capital punishment is vulnerable to overly simplistic interpretations. Despite these problems, I maintain that Kant has a good argument, one that is more compelling than we might expect. While Kant's bloody retributivist claims make many uncomfortable – and conflict with my own view that capital punishment is unjust – we cannot dismiss them as products of Kant's culture-bound preferences. (For an example of such a dismissal, see Pugsley 1981.²)

To this end, I will first argue that Kant's application of the *ius talionis* is consistent and coherent, and that it generates a prima facie requirement to execute murderers. Of course, the *ius talionis* generates a definitive requirement only if execution is a morally and legally permissible punishment. In the second half of the paper, I will show how Kant can establish the permissibility of execution by refuting the most powerful moral objections levelled against it. (I will briefly discuss the argument from fallibility in the conclusion.) The necessity of executing murderers follows from Kant's claim that there is a categorical requirement to impose the (permissible) punishments picked out by the *ius talionis*. So, by showing that execution is a permissible punishment, I show that it is, for Kant, a necessary one.

Since my aim is to show that Kant's justification is successful on Kant's own terms, I need to respond to the types of objections that could plausibly convince him, and *only* the types of objections that could do so. In this respect, the right to life and dignity objections are especially important because they try to turn Kant against himself: they invoke a Kantian categorical requirement to generate prohibitions on capital punishment. And the only objections that Kant can hear are those that issue from categorical requirements. This is because legal institutions are categorically required to punish (MM 6: 331), and hypothetical imperatives, such as those generated

by utilitarian considerations, cannot override categorical ones. Since this point about the categorical requirement to punish is crucial to all that follows, I will begin by explaining what it means.

Kant argues that, whenever people cannot avoid associating, they are categorically required to set up legal institutions (*MM* 6: 307). This is because legal institutions are necessary conditions of external freedom (*MM* 6: 306, 311; *TP* 8: 289): without legal institutions there would be no way of establishing, adjudicating and enforcing the rights claims that are the expressions of our external freedom.³ The categorical requirement to erect legal institutions emerges from that fact that, as rational agents, we necessarily value our freedom, and are thus committed to doing what it takes to actualize freedom in both its internal and external guises.

Now, the categorical requirement to establish legal institutions generates a categorical requirement to punish. This is because punishment is included in the concept of juridical law. In the 'Further Discussion of the Concept of the Right to Punish', Kant writes that 'the mere idea of a civil constitution among *human beings* carries with it the concept of punitive justice belonging to the supreme authority' (*MM* 6: 362). In other words, we cannot conceive of a lawgiving state without also conceiving of a state that punishes citizens for breaking the law. Kant argues for this claim earlier in the *Rechtslehre*, where he analyses the more general concept of practical lawgiving (*MM* 6: 218–21). What it means to give a law, he says, is to represent a specific action as necessary, to make sure that those subject to the law understand that they must act (or refrain from acting) in a certain type of way. But a lawgiver must also *motivate* those subject to the law to act in the way prescribed. The lawgiver accomplishes this by 'connecting' the representation of the law with an incentive to obey it (*MS* 218). Kant then draws a distinction between ethical and juridical lawgiving, a distinction which he bases on the mode of constraint belonging to each. Ethical lawgiving constrains action through the mere thought of the law, while juridical lawgiving constrains action through pathological incentives. That is, juridical lawgiving constrains action by exploiting our aversion to pain and attachment to pleasure, by threatening and meting out punishments. Kant sees this distinction as a necessary one: ethical laws constrain the ends we set for ourselves, and the pathological techniques of juridical law simply cannot force us to change our ends (*MM* 6: 218–21; 231). On the other hand, it is easy for legal

institutions to constrain actions through the threat and imposition of punishment. It is for this reason that punishment is part of the concept of juridical law – punishment is the mode of constraint that belongs to juridical law’s specific way of giving laws.

This conceptual point about punishment leads to a normative one. Since the practice of punishment is part of the concept of juridical law, and since the existence of juridical law, in the guise of legal institutions, is required by practical reason, the practice of punishment is required by practical reason. It is *categorically* required: ‘the law of punishment [*Strafgesetz*] is a categorical imperative’ (*MM* 6: 331).⁴ By this Kant means that punishment is not merely a justified response to lawbreaking that authorities may choose to apply at their whim. When a crime has been committed, legal officials must punish the wrongdoer – it would be unjust not to do so.⁵ Punishment is not merely a justified response to wrongdoing, it is a *necessary* one.⁶

Before moving to Kant’s justification of the death penalty, I would note that this account of the categoricity of punishment doubles as an account of Kant’s justification of punishment. Because juridical law is a condition of external freedom, practical reason categorically requires us to erect legal institutions. Because punishment, as the mode of constraint of juridical lawgiving, is contained within the concept of juridical law, practical reason categorically requires us (or, more precisely, legal officials) to punish. Punishment is justified because it is a necessary part of the conditions of external freedom.⁷

The coherence of Kant’s *prima facie* justification of capital punishment

Kant’s justification of the death penalty begins with the *ius talionis*. The *ius talionis* is the principle that specifies how legal institutions should determine the ‘kind and amount’ or ‘quality and quantity’ of punishment imposed on a lawbreaker (*MM* 6: 332). (The *ius talionis* is not, as sometimes thought, an instrument of vengeance; it is a rule for determining how legal institutions can hinder the lawbreaker’s freedom without hindering it too much and violating principles of right.) Kant clearly thinks that the *ius talionis* directs legal officials to execute convicted murderers (*MM* 6: 333). But Kant’s stated argument for this view is weak and open to attack.

Most importantly, he appears to use different conceptions of the *ius talionis* to determine the punishment of murderers and other wrongdoers, while failing to justify this difference. As such, Kant's claim that murderers ought to be executed can seem arbitrary and unmotivated. In the following pages, I will show how Kant can defend himself against this charge. In so doing, I will show how the *ius talionis* generates a prima facie justification of capital punishment.

Perhaps the most famous statement of the *ius talionis* is 'whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself' (MM 6: 332). I shall not be the first to note that Kant's discussion of the *ius talionis* does not deliver on the marvellous clarity promised by this quote. Kant's idea is that the relationship between crime and punishment must be one of equality (*Gleichheit*). But *Gleichheit* can mean, among other things, balance, similitude, uniformity, and consistency. Indeed, Kant seems to intend different senses in the *Metaphysics*. In some passages, the *Gleichheit* in question is a literal similarity between the deed that constitutes the crime and the deed that constitutes the punishment; for example, 'if he has committed murder, he must die' (MM 6: 333). I will call this 'strict retribution'. But, a page earlier, Kant offers a different interpretation of *Gleichheit*. There he describes penal justice as a balance, as 'inclin[ing] no more to one side than to the other' (MM 6: 332). Balance belongs to a proportional interpretation of the proper relation between crime and punishment. According to proportional retribution, punishment is just when the harm caused by the punishment is no less, and no more, severe than the harm caused by the crime. Kant appeals to the proportional interpretation when he says that someone who insults a social inferior should be made to apologize and kiss the inferior's hand (MS 332). Here the punishment imposed (kissing a hand) looks nothing like the criminal act (verbally insulting someone). But Kant's idea is that the wrongdoer is punished justly because he suffers roughly the same *amount* of harm that he imposed on his victim. By contrast, if the high-status wrongdoer were simply made to listen to a verbal insult, or pay a fine, he would not be punished severely enough, and his punishment would be unjust.

Although it is hard to measure harm in any precise way, and the *Gleichheit* involved in proportional retribution will therefore be less exact than that of strict retribution, Kant does believe that

proportional punishments can be a legitimate ‘substitute’ (*MM* 6: 333) for strict justice. (And, as we will see below, proportional punishments are sometimes a morally necessary substitute.) So it is clear that Kant endorses two different conceptions of equality in punishment, and two different conceptions of the *ius talionis*.⁸ Equality as similarity of deed generates a strict ‘eye for an eye’ version of retribution, while equality as similarity of harm generates a proportional version. It is the former conception of retribution that underlies Kant’s stated views on capital punishment. Kant claims that punishing authorities must *always* adhere to strict retribution when dealing with murders. For a murderer ‘there is no substitute that will satisfy justice’; ‘if he has committed murder, he must *die*’ (*MM* 6: 333). But the existence of two versions of retribution should lead us to question whether Kant has to make this claim, or whether his commitment to strict retribution is simply the product of arbitrary choice.⁹

I contend that Kant does have a good reason for thinking that murder must be punished under strict retribution, although it is not explicitly stated in the *Metaphysics*. I begin with Hill’s view that the *ius talionis* is basically a principle of proportionality, with the qualification that ‘the punishment ought to be, when morally and physically possible, the same kind of harm or loss the offender has wrongfully inflicted on his victim’ (1997: 302). Here the idea is that, when strict retribution is both possible and permissible, legal officials do not need to measure the severity of punishments and match severe punishments to severe crimes. When legal officials impose the same *kind* of harm on the offender that the offender imposed on his or her victim, punishment is at its most exact. (Morally possible punishments are punishments that do not violate the dignity of the offender. Legal institutions cannot rape rapists because doing so would violate the lawbreaker’s dignity and would be a moral wrong, as well as a ‘punishable crime against humanity as such’ [*MM* 6: 363]. Physically possible punishments are just that; it is impossible to apply strict retribution to a blind person who blinds someone else.)

Kant thinks legal institutions ought to use a strict retributive calculus when possible for the same reasons legal institutions ought to use a retributive calculus at all: retributive procedures for determining punishment are the most resistant to all the ‘extraneous considerations’ that can enter into sentencing and render it unjust

(MM 6: 332). Legal institutions ought to exclude these factors, because punishment is just only in so far as it accords with laws of freedom, and laws of freedom require that legal institutions treat all citizens, and criminals, as ends in themselves (MM 6: 331). Choosing punishments based on maximizing deterrence, satisfying vengeance or compensating victims, would thus be unjust.

So, unless the punishment picked out by strict retribution is impossible or immoral, it ought to be imposed. In fact, the more severe the crime, the more pressure there is to impose punishment in accordance with the strict *ius talionis*. Since the amount of punishment, the amount of hindrance of an offender's freedom, will increase with the severity of crimes, the injustice of an incorrect sentence increases as the severity of the punishment increases. And Kant thinks the best way to avoid injustice is to impose punishment in accordance with strict equality.

If this interpretation of the *ius talionis* is correct, showing that murder is one of the worst (or most severe) crimes will provide a prima facie reason for using strict retribution and executing murderers. Common sense often says that murder is one of the worst crimes because it destroys a life, and life is the most valuable thing that human beings possess. We find something like this view in Kant. Life, he says, is both the capacity to represent things as desirable and the capacity to accomplish that which is desired (MM 6: 211). Since life is the condition of achieving happiness, assessments of the value of life made from the perspective of happiness will rate life higher than anything else.

But we need to be careful in specifying why life is valuable. We can get misled about the value of life if our judgments are swayed by an attachment to happiness. In Kant's view, human life is valuable because it is the life of a rational being; to put it crudely, human life is valuable because it is the vehicle we use to exercise our freedom. Murder, then, is one of the worst crimes because, by destroying a life, a murderer destroys a particular instantiation of what has absolute value. (More specifically, murder destroys the body, which is the 'total condition' of life for human beings [Col 27: 369].) While Kant does not explicitly characterize murder in this way, he does say that killing *oneself* is wrong because it 'annihilate[s] the subject of morality in one's own person', which is, to a limited extent ('as far as one is capable'), 'to root out the existence of morality itself from the world' (MM 6: 421–2). The same point can be made with

regards to murdering others. Because murder aims at rooting out the existence of morality itself, it is not just different in degree from other crimes, it is different in kind. I will discuss this in more detail below, but even this brief sketch should show that murder is a *sui generis* crime.

So Kant *does* have a coherent retributivist argument for the execution of murderers. The prospect of severe punishment pressures authorities to punish in accordance with strict retribution. Since taking a life is one of the worst possible crimes, it will be punished very severely, and thus legal institutions cannot risk the flexibility and judgment involved in proportional punishment. If murder is one of the worst possible crimes, it is important to treat it as justly as possible, and this requires the use of the strict version of the *ius talionis*.

Although Kant's explicit endorsement of capital punishment makes use of strict retributivist considerations, we should not exclude the possibility that execution could be attached to crimes through proportional retribution. The basic idea of Kant's proportional retribution is that both crimes and punishments are ranked in severity, and that punishments are attached to the crimes to which they are equal in severity. (Kant never states this explicitly but, as Murphy notes [1987: 532], Kant probably thought it was self-evident). For example, if we ranked punishments and crimes on a scale of 1–10, crimes that were an 8 would merit punishment that was an 8. A legal institution that punished a crime of 8 with a punishment of 8 would be punishing properly, and one that punished the same crime with a more or less severe punishment would be punishing improperly.

Of course, it is not enough to attach two ordinal scales to each other. There must be some substantive correspondence (which philosophers of punishment call 'commensurability') between the two scales of severity. If a state's most severe punishment were ten days in prison, this would be an inappropriate punishment for the worst type of wrongdoing. Kant does not tell us how to measure this correspondence, but I would follow Hill in suggesting that the scale of punishments should be determined by how rational people, from a self-interested point of view, would rank the undesirability of all the types of permissible punishments (1997: 299). The severity of crimes should be ranked in a similar way – by thinking about how a rational person would rate the hindrance to freedom created by various offences.¹⁰ (Kant has no illusions that constructing a

comprehensive ranking would be an arduous, if not impossible, task, and this is another reason for using the strict *ius talionis* when possible.¹¹⁾

I have already explained why murder is one of the worst crimes. Kant thinks it obvious that the most severe punishment – the most severe morally permissible punishment, as I will explain later – is execution. In the *Anthropology* he claims that love of life is a human being's strongest impulse (A 7: 276). Since execution takes away one's life, execution is the greatest harm one can imagine, and threats of execution will generate the greatest amount of aversion. (Prolonged torture may be worse than execution, but torture is impermissible.) So execution, as the worst punishment, would be attached to murder, the worst crime.

This proportional account helps us understand why Kant thinks execution is an appropriate punishment for rebellion (MM 6: 320). Rebellion is a crime for which punishment in accordance with strict retribution is impossible. But rebellion is like murder in the wrong it does. Rebellion and murder share a feature that ranks them as the worst of all possible crimes: both attack the condition of possibility of freedom. Murder eliminates an individual's capacity to exercise her internal or external freedom. Rebellion, when successful, destroys the institution that is designed to enhance and ensure external freedom. Since rebellion is in the class of the worst crimes, proportional retribution would punish it with execution. That said, I am not convinced that Kant's justification of capital punishment can legitimize the execution of rebels, for reasons explained below. As a result, the rest of the paper will focus on the execution of murderers.

There is a potential complication with my defence of Kant's claim that the *ius talionis* directs legal officials to execute murderers. One might question whether Kant has a reason to accept the *ius talionis* as a standard of punishment in the first place. The justification of punishment advanced above does not obviously require legal institutions to punish in accordance with the *ius talionis*.¹² And, if the *ius talionis* is an unnecessary feature of Kant's theory of punishment, his justification of the death penalty will fall apart.

Kant seems to think that the categoriality of punishment leads to the necessity of the *ius talionis*. But his argument for this view, such as it is, is unconvincing. Kant rightly notes that the categorial imperative of punishment forbids states to refrain from punishment

in order to achieve happiness on the part of society or the wrongdoer. Considerations of happiness, and related ends, must not influence *whether* states punish. Kant can naturally extend this point to say that considerations of happiness cannot influence *how* the kind and amount of punishment should be determined. In this way, categorical considerations lead to a rejection of classical utilitarian justifications of punishment, as well as a rejection of classical utilitarian formulas for determining the kind and amount of punishment (see, for example, *MM* 6: 331). But, as Merle points out, there is a problem here. Unless deterrence and retribution are the only two conceivable theories of punishment, Kant's argument does not show that the *ius talionis* is the only acceptable means of determining punishments (Merle 2000). It shows only that utilitarian considerations are unacceptable means of determining punishments.

While this problem one is an interesting one, it need not detain us at length. Principles of justification rarely generate comprehensive and satisfactory policies of application.¹³ Kant clearly thought the *ius talionis* was the punishment policy that fitted most comfortably with his general practical and conceptual commitments, and it is not hard to see why. Kant's choice of the *ius talionis* responds to the concern that criminals be treated with the respect due to them as human beings. Treating them with respect means treating them equally, and making sure that 'extraneous considerations' irrelevant to the assessment of the wrongdoing do not interfere with sentencing. And the best way of making sure that 'extraneous considerations' do not affect sentencing is, as I discussed earlier, to strive for an exact equivalence between the amount of freedom taken by the wrongdoer and the amount of freedom taken from the wrongdoer.¹⁴

So at this point in the paper I have shown that Kant's theory of retribution is consistent, and that it directs legal officials to impose the death penalty on murderers. Now it is time to deal with the more difficult part of his justification. Given Kant's insistence that moral norms place restrictions on punishment, the scale of punishments will have an upper threshold marking punishments that are impermissible.¹⁵ The argument I have been attributing to Kant presumes that execution is a member of the class of permissible punishments, or punishments that are lower in severity than the upper threshold. It presumes that executing a murderer is permissible in a way that raping a rapist is not. But this must be defended.

I will show how Kant can defend this presumption by showing how he can respond to the two most important arguments against his position on capital punishment. These arguments are potentially damaging, because they use a Kantian categorical requirement to generate a prohibition on capital punishment. As I stated above, I focus on this kind of objection because my aim is to show that Kant's justification is successful given his presuppositions. And, since the law of punishment is a categorical imperative, the only type of objection Kant can hear is one that issues from a similar type of imperative. (That said, the objections considered here are probably the most popular ones, as well.) The first objection states that, since the right to life is inviolable, the state cannot put even a criminal to death. The second builds on Kant's view that the state cannot violate the dignity of even the most grievous wrongdoer, adding the idea that execution is just such a violation.

The right to life

In the context of arguments against the death penalty, the 'right to life' refers to a duty on the part of others not to take our lives. Although Kant never uses right-to-life language, he would certainly agree that a citizen has a right to life in the sense that (a) other citizens have a duty not to take his life, and (b) legal institutions are authorized to coerce citizens not to take each other's lives. But this uncontroversial point of agreement is not enough to support a Kantian argument against capital punishment, as it says nothing about whether this right puts a constraint on legal institutions. I can see two ways of asserting that there is such a constraint – one juridical, the other moral – and I will discuss them in turn.

An opponent of capital punishment might argue that there is something contradictory about executing a murderer. For Kant, the whole point of legal institutions is to secure and enhance external freedom. Since executing a murderer eliminates one more instantiation of freedom, the argument would go, it runs counter to law's purpose. This juridical objection has two problems. First, it proves too much. For Kant, all punishment infringes on freedom; to infringe on freedom is just what it means to punish. If we accept the objection, we have to reject Kant's theory of punishment as a whole. In fact, if we accept the objection, we probably have to reject an

entire family of justifications of punishment. Under almost any non-consequentialist account, punishment will have adverse effects on the wrongdoer, thereby impinging on whatever values law is meant to protect. So the argument on offer is really an argument against punishment as a whole. And the vast majority of critics of capital punishment, myself included, are ill-disposed toward pressing this line of attack.

Second, the juridical objection misstates what Kant thinks law is meant to do. The point of law is not to enhance *criminals'* freedom, the point of law is to enhance *citizens'* freedom. A criminal forfeits his right to his external freedom (TP 8: 292).¹⁶ He gives up his civil personality, and along with it, his civil independence, the quality of 'owing his existence and preservation to his own rights' (MM 6: 314). As such, a criminal loses his status as a rights-bearing agent, and he loses his status as an object of the state's protection. If a criminal has no juridical rights, he certainly has no juridical right to life.

Now for Kant, criminals do retain 'rights' in the sense that legal institutions cannot treat criminals in any way they please. Criminals cannot be treated like things (MM 6: 463–4). While criminals can forfeit their civil personality, they can never forfeit their innate personality (MM 6: 331).¹⁷ This reference to innate personality shows that there is room for appeals to a right to life in *moral* terms. And we should probably take the moral appeal more seriously. When we speak of the rights Kant would grant a criminal, we are really speaking of the practical constraints generated by the criminal's status as a being with personality.

At their most basic, moral right-to-life arguments proceed from the familiar prohibition against treating people as mere means to an end (MM 6: 331, 462). Since the prohibition against treating people as mere means is a categorical one, the moral approach appears to show that our right to life – the duty on the part of others not to kill us – is absolute. After all, taking someone's life destroys an end-in-itself. Killing a person turns a moral agent into a corpse-object. If reducing a moral agent to thinghood does not violate the prohibition against treating people as mere means, it is hard to see what would. So the Kantian critic of capital punishment will argue that, since execution eliminates an end-in-itself, execution violates the moral law (Pugsley 1981; Schwarzschild 1985; Merle 2000).

Put this way, the objection sounds less like a claim about rights and more like a claim about dignity. Indeed, critics who level the moral right-to-life objection typically consider their arguments to be about human dignity.¹⁸ I distinguish right-to-life from dignity objections because there is an important difference between them. Let me begin with their similarity: both objections make use of the absolute value expressed by our status as beings with dignity. Dignity is a complex concept, but for our purposes it is enough to say that dignity names the elevated status held by autonomous beings (see, for example, G 436). Now here is the difference: *moral* right-to-life objections work by transferring the value that accrues to our existence as dignified beings to life. Dignity objections remain tightly focused on the concept of dignity itself. While this might seem to be a mere difference in emphasis, right-to-life objections go wrong in a different way from dignity arguments.

Moral right-to-life objections go wrong because the transfer of value is illegitimate (I will explain how dignity arguments go wrong in the next section). While our existence *qua* free beings counts as an end-in-itself, it does not follow that our *lives* are ends-in-themselves. That we are beings with dignity does not mean that our lives have a value beyond all price. Kant makes this clear in some of his descriptions of what it means to act freely. In the *Religion*, he describes a person who actualizes his freedom as a being who is ‘capable of possessing and adopting as his goal something [viz., moral freedom or autonomy] . . . which he values more highly still than his life, and [of] sacrificing all self-interest to it’ (*Rel* 33n, emphasis mine). Kant provides an illustration in the *Critique of Practical Reason*, where he explores the hypothetical case of a man whose prince has commanded him to give false testimony against a neighbour on the pain of execution. This man, Kant says:

would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. (*CPrR* 30)

Free beings do not consider their lives to have absolute value, and they recognize that there are cases where they must give up their lives in the service of moral ends. (That the man is exhibiting freedom is shown by the context – a discussion of the Fact of Reason meant

to illustrate what the consciousness of subjection to the moral law might look like.) The point here is that, while our dignity confers a high value on life, acting in accordance with dignity has an even higher value. So, while life is a vehicle for a being to express its freedom, life as such is not an end-in-itself. Life has value only in so far as it serves our moral existence. Sometimes, as I will discuss in the next section, the preservation of our moral existence depends on the elimination of our physical existence.

In short, the value of life is a relative one; ‘it is better’, Kant writes, ‘to sacrifice life than to forfeit morality’ (Col 27: 373). Demonstrating the relativity of the value of life defeats the moral right-to-life objection in the following sense. If life does not have absolute value, and is not an end-in-itself, there is no categorical prohibition on the taking of life, and no absolute right to life. Showing that there is no absolute right to life shows that the death penalty cannot be criticized for violating this alleged right.

There still might be room for an objection. Moral right-to-life objections can work even if there is no *inviolable* right to life. As a vehicle for expressing freedom, life still has significant value, enough to generate what Hill calls a Kantian presumption against the state’s taking the lives of its citizens (2003: 26, 31). It is unclear how much weight Kant would give to this presumption. My discussion of the *ius talionis* provides good reason for thinking that requirements of justice outweigh this presumption, but I am not sure it settles the matter. I will offer a conclusive defeat of the right-to-life argument at the end of the next section.

Death before dishonour

Now I want to turn to the objection that execution violates the wrongdoer’s dignity. An act violates dignity when it treats a free being in a way that does not accord with that status. So punishment violates dignity when it treats a free being in a way that does not accord with that status. Kant rejects drawing and quartering as impermissible for this reason (*MM* 6: 463). The question for us is whether execution treats someone in such a way. There is nothing in the drawing and quartering example, nor in the *Tugendlehre*’s discussion of violations of dignity, that decides the question with respect to execution. In my view, the best way to figure out whether

a type of punishment violates dignity is to determine whether a person proper could will that punishment to be imposed on herself. Since a person proper wills in accordance with practical laws, a person proper will not will anything that contravenes their dignity. So, if one shows that a person could will a type of punishment, one shows that the type of punishment under consideration does not violate dignity.¹⁹

I will argue that a person could will her own execution. By showing that it is possible for a person to will her own death at the hands of the law, I will show that the dignity objection fails. (I do not think that we can determine what sorts of punishments Kant thinks persons would will to be imposed on themselves in all cases. However, we can know what sort of punishment a person would will to be imposed on herself if she murdered. This is because Kant discusses in detail what the act of murder says about a murderer.)

The starting point for my argument is a claim Kant makes about an honourable person's views about execution.²⁰ In the *Rechtslehre*, Kant asks us to consider a scoundrel and an honourable man, both of whom have committed a capital crime. (Kant uses rebellion as the crime in this example, but I will substitute murder.) Kant claims that, if each were given the choice of execution or a life of convict labour, the honourable man would choose death and the scoundrel would choose life (*MM* 6: 333).²¹ By examining the honourable man's choice, we will see why a person proper would will his own execution.

It is important to note that, when Kant discusses the honourable man's choice, he is referring to the a priori, or anthropomorphical, sense of honour, not honour in its various empirical, or anthropological, guises (*MM* 6: 406). This is an important distinction, because some of the empirical manifestations of honour – such as military honour, which motivates insulted soldiers to kill their fellows in duels (*MM* 6: 336–7) – can lead us morally astray.²² In fact, Kant thinks that moral wrongdoing is often motivated by the desire to be honoured by others (*A* 7: 271–3). But, in the anthropomorphical sense, the desire for honour is equivalent to the desire to be able to respect and value oneself; it is the desire to ensure that one's actions and dispositions are *objectively* worthy of respect (*Col* 27: 347; *A* 7: 272). The difference is that the morally valuable desire for honour aims at self-esteem in relation to the moral law, while the morally indeterminate desire for honour aims merely at esteem

in the eyes of others, an esteem that can be granted for good or bad reasons.²³ Now, in so far as pure love of honour refers to a virtuous way of thinking that inclines people to do what duty commands (MM 6: 420), the similarity between love of honour and respect for the moral law is very strong. As a result, the fact that Kant has the honourable man choose execution indicates that this choice is morally sound.

Before seeing why that might be, I would note that, *if* execution is permissible, the honourable man would clearly choose it. Kantian retribution assigns execution to the man's crime and, since the honourable man wants to act in accordance with practical laws, he will want justice to be done. This means that he will choose execution over a lesser punishment, since execution is required by the *ius talionis*. But we must set this point aside, as we are interested in an explanation of the permissibility of execution that does not beg the question. My explanation hinges on the fact that murder puts a stain on the perpetrator's honour that cannot be removed. By continuing to live with this stain, rather than be put to death, the honourable man is *dishonoured*.

All wrongdoing stains dignity to some degree, but murder is qualitatively different from other crimes. Murder is as close as someone can come to pure evil, which Kant defines as rejecting the authority of moral law as such (MM 6: 320–2n).²⁴ Kant thinks that it is impossible to be purely evil (MM 6: 320–2n); since the moral law has authority over all free beings and only free beings, to reject the authority of law would be to reject one's own freedom. And *this* is impossible for any number of reasons; to take just one example, any rejection of freedom would be an act of freedom. Now, murder is not purely evil – murderers do not reject the authority of law, and they do not deny the value of freedom as such. (If a capital defence attorney can show that her client believes that murder is always the right thing to do, the attorney will argue that her client is insane, and not someone to whom culpability can be attributed.) While murderers do deny the value of their victim's freedom, they do not deny the value of their *own* freedom. Murderers typically do not want to be murdered themselves, and one can imagine that most murderers would complain of wrongdoing if anyone tried to murder them. Nevertheless, I will argue that murder is close to pure evil because it enacts what we might call a *local* rejection of the moral law.

In Kant's view, most, if not all, types of interpersonal crimes involve an illegitimate manipulation of someone's freedom. But, for all crimes other than murder, this manipulation is limited in its scope. If I lie to someone about the location of the nearest dry cleaner, coercing her into going somewhere she does not want to go, my manipulation lasts a short period of time, and affects my victim's freedom only to a limited extent. She still has the capacity to buy an ice cream at the store she discovers along her way. Murder, on the other hand, achieves a 'mastery' over its victim (Col 27: 372). Murder eliminates freedom with a finality that other crimes do not achieve; as far as we know, death does not end, and when you're dead, you're dead. To be sure, we can imagine some heinous crimes whose effects never cease – a victim of a particularly vicious assault might never fully recover, physically or emotionally. But this victim still has avenues for expressing her freedom, even if they are severely curtailed. Murder robs its victim of any such avenues. So, while the murderer does not renounce the value of freedom as such, he does deny – in a very thoroughgoing sense – the value of one instantiation of freedom. He denies that the victim's freedom exerts even the most minimal claim to his respect. In this way, a murderer enacts a local rejection of the authority of law.

It is this feature of murder that drives the honourable murderer's choice. He realizes that he has done something that constitutes the most egregious violation of the moral law possible. (He also realizes that he has the capacity to repeat this offence in the future. If he continued to live, there would be a constant threat that he would act on this capacity. This prospect would be intolerable.) Since, as an honourable man, he considers life worth living only as an honourable life, he would welcome the state removing the stain on his dignity by taking his life, especially because he is forbidden to commit suicide. In short, the honourable murderer chooses death in order to preserve his dignity. I want to stress that this is not merely a psychological description about what honourable people in fact do; it is a claim about what one ought to do. '[In] the cases where a man is liable to dishonour', Kant writes, 'he is duty bound to give up his life, rather than dishonour the humanity in his own person' (Col 27: 377).

This mention of suicide should not lead one to worry that the honourable man's choice violates the duty of self-preservation. This duty is not absolute: 'to preserve my life', Kant writes, 'is only a

conditional duty’ (TP 8: 300n, emphasis mine; see also Col 27: 371). In the *Rechtslehre*, he says that citizens have a duty to obey the state when it requires them to risk their lives on the battlefield (*MM* 6: 345–6; see also Col 27: 371). He elsewhere discusses an example of two survivors of a shipwreck floating on a piece of wreckage that will presumably not hold both of them for long. It is wrong, he asserts, for one survivor to push the other into the sea in order to save his own life (TP 8:300n). The conditionality of the duty of self-preservation should be no surprise, given Kant’s conception of the value of life. Since life is valuable because it is a vehicle for moral existence, morality will always have a priority over self-preservation. Kant asserts this priority quite forcefully:

It is better to sacrifice life than to forfeit morality. It is not necessary to live, but it is necessary that, so long as we live, we do so honourably; but he who can no longer live honourably is no longer worthy to live at all. (Col 27:373)

Living dishonourably extends our biological life at the cost of our ‘moral life’ (Col 27: 377). Faced with a choice between living dishonourably and preserving one’s biological life, one should give up one’s biological life, since it has no intrinsic worth and is not an end-in-itself.²⁵ And this is just the choice faced by the honourable man.

Since any person acting freely would make the same choice as the honourable man, the honourable man’s response should be generalized to persons as such. That is, those who commit murder, would, as beings with personality, choose that they, *qua* criminal, be executed.²⁶ So Kant has a response to the claim that execution violates human dignity. By showing that persons could, indeed *would*, will to be executed, were they to murder, he shows that there is no such violation.

These considerations also show that Kant can rebut the right-to-life argument left over from the previous section. They show that the presumption against the state’s taking the life of a citizen does not bar execution. The presumption is grounded in the value of dignity, but a citizen’s life is worth only the dignity to which it is laminated and, in the case of murderers, that dignity can be preserved only at the cost of the offender’s life. So the presumption is outweighed by the fact that refraining from executing a murderer would degrade her, violating the very value the presumption is

meant to protect. There are at least two ways in which refraining from execution would degrade the murderer. First, by keeping her alive, the state would preserve her in the condition of dishonour. The state would be refusing to provide the one option she has for giving up her life in accordance with practical laws. Second, the state would be refusing to treat the offender as she would rationally will to be treated. In this sense, the state would be treating her not as a rational end-in-herself, but as a child or an animal.

Conclusion

This completes my account of Kant's justification of the death penalty. Before moving to my final remarks, I want to state the argument in its entirety. Kant's theory of the *ius talionis* provisionally attaches execution to murder and rebellion. The gravity of these crimes triggers the need to punish them in accordance with the strict version of the *ius talionis*. Since the law of punishment is a categorical imperative, legal authorities must impose the punishment required by the *ius talionis*, unless (Kantian) moral considerations prohibit it. And, as I have shown in the previous two sections, the most obvious and important moral considerations – the right to life and the inviolability of dignity – do not prohibit capital punishment, at least for murder. Therefore, legal institutions are required to execute murderers whose crimes can be imputed to them. (I add this last qualification because Kant defines crime as an intentional transgression and defines a criminal as someone who is culpable for his transgression [MM 6: 224]. These definitions preclude states from executing minors, as well as the insane. They also generate distinctions between degrees of criminal homicide that would cash out in different punishments for homicides of different degrees. For example, homicides resulting from criminal negligence involve less culpability than premeditated murder, and would therefore be punished less harshly.)

Although I have argued that Kant's justification of capital punishment is coherent, I do not think that it sanctions the execution of rebels, despite Kant's stated views on the matter. Kant equated rebellion and murder because they both attack the condition of the exercise of freedom. However, while rebellion may, at first glance, look as if it shares the *sui generis* status of murder, it does not. First,

a legal institution can punish rebels only when the rebellion fails. And, if the rebellion fails, there is no elimination of the condition of (external) freedom. An attempted rebellion might attack the condition of external freedom, but in that sense it is no different from any other crime. Second, even successful rebellions do not accomplish an irrevocable destruction of freedom, as most rebellious regimes install a legal system sooner or later. Rebellions might not even temporarily detract from citizens' capacity to exercise their external freedom. 'Good' rebellions can even strengthen that capacity. So I think it is an open question, at the very least, whether execution is a permissible punishment for rebellion.

To conclude, I want to address the third objection to capital punishment mentioned earlier. It is not strictly Kantian, but quite popular in the abolitionist literature. This objection grants that legal institutions may be authorized to execute, yet asserts that they ought not to, given human fallibility and the procedural imperfections that follow (see, for example, Nathanson 2001). I have suggested that, the more severe the punishment, the more important it is that sentencers get the punishment right. The US Supreme Court expresses this view in the 'death is different' doctrine, holding that there is a need for 'heightened reliability' in capital sentencing (*Woodson v. North Carolina*, 1976). This means, in part, that sentencers must adhere to special procedural safeguards designed to mitigate against the arbitrary and/or capricious imposition of death sentences (*Gregg v. Georgia*, 1976). However, there is compelling sociological evidence that execution is disproportionately meted out to those who murder whites as opposed to those who murder African-Americans (Baldus et al. 1983). This is only one of many well-documented procedural problems with the current system of capital punishment in the United States, but it is enough to illustrate the objection – if legal institutions cannot eliminate racial considerations from sentencing, the death penalty is unjust in application, whatever its merits in theory.

I think Kant would consider race to be one of the 'extraneous considerations' that ought to be excluded from determinations of punishment. But legal sociology did not exist in his day, and it is hard to know what practical conclusions he would draw from an examination of these systemic inequalities. This issue is compounded by the fact that Kant, so careful to account for finitude and fallibility in his other major works, neglects to do so in the *Metaphysics*. The

Metaphysics is concerned with the *idea* of a civil society, the ideal of civil society, the 'form of a state as such' (*MM* 6: 313). The picture of legal institutions that we find in the *Metaphysics* is a regulative idea, and thus presents a conception of legitimacy that actual legal institutions can never reach. And Kant makes no provisions for dealing with the imperfections that plague actual legal institutions, except to mention in passing that the state should try to improve its laws (*MM* 6: 372). (Nor does he ever really address the question of what people ought to do when they live in the shadow of a corrupt or repressive legal institution.)

Some scholars have argued that it is possible to construct a Kantian non-ideal political or legal theory, and I think the task is feasible (Korsgaard 1996: 147–54; Holtman 1997: 19–21). However, I am not sure such a task is needed in order to rebut the objection on offer. This is because the fallibility objection has important limitations. In a sense, it does not reach the death penalty itself. It is a proceduralist argument, and a state could answer it by diverting more resources to improve the reliability of capital sentencing. Even if such reform were enacted – which it certainly ought to be – there would be certain cases, such as when a murder was witnessed by multiple people of unimpeachable credibility, where the death penalty would go through. At its most successful, the objection might generate an admission that there ought to be a moratorium while capital sentencing is reformed, but the objection cannot tell us what is wrong with capital punishment specifically.²⁷

Notes

- ¹ I take this phrase from Murphy (1987: 521).
- ² It should be noted that, in Kant's day, execution was accepted as an appropriate punishment for all manner of minor crimes (Potter 2002: 272–3). In this respect Kant was, as Potter notes, more progressive than most of his contemporaries.
- ³ All practical laws, in the Kantian sense, have freedom as their end, however indirectly (*MM* 6: 221). But the type of freedom aimed at by basic juridical principles is different from that aimed at by basic ethical principles. Legal institutions aim at securing and enhancing our external freedom, our freedom to live our lives in accordance with the permissible ends we set for ourselves absent illegitimate interference by others. Unfortunately, Kant is vague about the differences between internal and external freedom. For a helpful discussion, see Uleman (2004).

- ⁴ Kant says that, if a state has to execute so many people that the state would dissolve and regress to a state of nature, the state must forgo execution (*MM* 6: 334). This complication is not, as some would have it, evidence of Kant's deterrent tendencies. Rather, Kant is simply pointing to the consequences of the fact that punishment gets its categorical status from the way it secures external freedom. If a certain application of state punishment would eliminate citizens' external freedom – as it would in this special case – that punishment would no longer be categorically required. Indeed, it would be prohibited.
- ⁵ This way of explaining the legitimacy of punishment can be rephrased in terms of the famous hindrance principle – Kant's claim that hindering hindrances to freedom promotes freedom (*MM* 6: 231). The hindrance principle is Kant's generic way of showing that an authorization to coerce is part of the concept of right; the hindrance principle states that this connection holds in the state of nature as well as civil society. The argument I have provided draws on Kant's more specific conception of the role of juridical law in civil society.
- ⁶ It may be possible that some rational beings inclined to disobey the law would do better with a different kind of external constraint. But we must remember that the *Metaphysics* is designed to take into account the 'particular nature of human beings' (*MM* 6: 217). And criminal law, Kant thinks, is just what human beings need.
- ⁷ Most recent work in the Anglo-American world has rejected the traditional picture of Kant as a 'thoroughgoing retributivist' in favour of one in which he holds a mixed theory of punishment (Scheid 1983; Byrd 1989; Holtman 1997). According to the mixed theory, Kant justifies the basic institution of punishment by reference to its deterrent, or crime-preventing, effects, and uses retributive principles to determine the kind and amount of punishment that legal institutions can legitimately impose. On this interpretation, retributive principles generate side-constraints that limit how punishment can be carried out.

My interpretation of Kant's justification of punishment modifies this view in important ways. While advocates of a mixed theory are right to deny that Kant is a thoroughgoing retributivist, and their interpretation is more faithful to the nuances of Kant's theory of punishment, there are serious problems with describing Kant's basic justification of punishment in terms of deterrence. Most importantly, this description fails to account for the categorical nature of punishment. To see this, we can turn to Byrd, who offers the most developed defence of the mixed theory. Byrd claims that the 'the purpose of the criminal law is to protect this social order' (1989: 154). In her view, it is by appeal to this purpose that criminal law is justified.

The problem is that Byrd's conception of the purpose of criminal law fails to capture the tight relation between the value of freedom and the purpose of punishment. As a result, it leaves the idea of protecting the social order dangerously underdetermined. Her interpretation can easily lead to a full-blown deterrence theory of punishment, whereby protecting the social order is the ultimate aim of punishment. And if deterrence

is the basic justification of punishment, it is difficult to make sense of Kant's view that punishment is a categorical imperative. This is because deterrence theories can posit ends for criminal law that have no categorical value. My own formulation would be: 'the purpose of the criminal law is to enhance and extend external *freedom*.' This formulation captures the relation between the value of freedom and the purpose of punishment, preserving the categoricity of punishment.

- ⁸ The distinction between strict and proportional retribution is not entirely crisp. It is easy to turn proportional retribution into strict retribution by changing our description of the acts involved. So, in the example discussed above, both the crime and the punishment could be considered acts of humiliation. And, at a very general level of act description, all crimes are infringements on freedom, and all punishments are infringements on freedom. So, at the general level, all punishments can be viewed as strictly retributive, even though differences between crimes and punishments emerge at more specific levels of act description. This point can help explain why Kant describes the punishment for theft in ways that strike us as strictly retributive, even though the punishment is clearly proportional (*MS 6: 333*). Either Kant was confused about the level of act description under consideration, or he was sloppy in identifying it.
- ⁹ Sarver discusses this point at length (1997). He thinks the two versions are incompatible, and he argues that Kant must adopt the proportional version, since this version is more generally applicable to crimes. For Sarver this means that the *ius talionis*, properly understood, does not require the imposition of capital punishment. On the interpretation of the *ius talionis* I defend, the two versions are not incompatible, and Kant's retributivist argument is vindicated on this count.
- ¹⁰ Hill says that crimes should be ranked by thinking about how a rational person would rank the harm or loss incurred by herself as a victim of crime. I do not think this is quite right, since, as we see in Kant's discussion of theft, crimes are, in part, offences against society at large (*MM 6: 333*). Furthermore, thinking of harm in Kant's way will make it much easier to attach punishments to crimes that have no victim, as well as to criminal attempts.
- ¹¹ For more on the difficulties involved, see Russ Schafer-Landau's 'Retributivism and Desert' (2000) and Claire Finkelstein's 'Death and Retribution' (2002).
- ¹² Kant's use of this standard does not even follow from all thoroughgoing retributivist justifications of punishment. For example, 'evil deeds deserve punishment' does not state that punishment must be meted out in accordance with the *ius talionis*. The *ius talionis* would more easily follow from a moral principle such as 'evil deeds deserve punishment equal in severity to the evil of the deed'.
- ¹³ I take the distinction between principle and policy from Hill (1997: 310).
- ¹⁴ This interpretation is by no means novel. It is shared by Byrd (1989: 196–7), Hill (1997: 310) and Clark (2004: 16), among others.

- ¹⁵ Retributivist theories also need to set a lower threshold (on the crime side) marking infringements of freedom that do not merit punishment. Kant does not have much to say about this. Höffe claims that Kant thinks of crimes as ‘serious forms of violations’ such as felonies, as opposed to misdemeanours and infractions (2002: 157).
- ¹⁶ Kant never explains why a criminal forfeits her rights, and is somewhat vulnerable on this point. In my view, the forfeiture of rights follows from the state’s authorization to punish. If we accept that authorization, we must accept that the criminal has no right to coerce legal authorities not to infringe on his freedom. If the criminal had such a right, the state would not be able to secure the freedom of its citizens. And that is just what the state must do.
- ¹⁷ Kant does say in one passage that criminals can be a ‘mere tool’ of another’s choice (*MM* 6: 330). However, textual evidence massively favours the view that a criminal must be treated as a being with dignity.
- ¹⁸ This may be to avoid obvious difficulties of right-to-life arguments. For example, to claim that the right to life is absolute is to say that it is always morally wrong to send people to war, or to kill in self-defence, and it is difficult to defend either of these claims. It is for this reason that Pugsley explicitly rejects right-to-life arguments (1981: 1521).
- ¹⁹ I am not suggesting that types of punishment accord with dignity only when a lawbreaker consents to be subject to them. Kant is right to say that this would be absurd, as it would base the authorization of punishment in the lawbreaker’s choice (*MM* 6: 335). What one must look at is what a person *qua* homo noumenon could will for herself, *qua* homo phenomenon, were she to commit a crime. One could also employ the *Wille/Willkür* distinction to make this point, although Kant did not avail himself of this strategy.
- ²⁰ Honour has received little attention from Kant scholars. Susan Shell and David Sussman are the exceptions (Shell 1997; Sussman 2008); Tom Sorell makes some brief remarks about honour (1988). I want to thank an anonymous reviewer for drawing my attention to Sussman’s paper.
- ²¹ In his lectures, Kant generalizes this comparison: ‘There are duties that are far higher than life, and that must often be performed by sacrificing life . . . A man of inner worth will sooner sacrifice his life than commit a disreputable act; so he puts the worth of his person above his life. But the man without inner worth would sooner commit a disreputable act than sacrifice his life’ (*Col* 27: 375–6).
- ²² For more on Kant’s ambivalent views on degraded forms of honour, see Sussman (2008).
- ²³ This way of putting the distinction does not always hold. In at least one case, Kant says that pure love of honour aims at establishing our dignity in the eyes of others (*MM* 6: 464). But even here, it is not others’ esteem that is the true object of desire. What the honourable person desires is that he or she have as much dignity as his or her fellows.

²⁴ Actually, the formal execution of a monarch is even closer, but this special case does not affect my argument (*MM* 6: 320–2n).

²⁵ These considerations differentiate the giving up of life from suicide. Suicide 'dispos[es] of oneself as a mere means to some discretionary end' (*MM* 6: 423). That discretionary end is, of course, happiness. Someone who commits suicide kills herself in order to accomplish pleasure, by ridding herself of all her troubles (*Col* 27: 343). But, in the examples discussed above, sacrificing one's life is not a means to a discretionary end; it is instead a response to a demand of pure practical reason; 'to sacrifice one's life', Kant writes, 'is not suicide' (*Col* 27: 371). Furthermore, the requirement to choose death over dishonour obtains only when we do not need to proactively take our own lives; Kant says that giving up our lives is required by the moral law only when we are not the 'agent' of our deaths (*Vig* 27: 629). This qualification seriously limits the types of cases where sacrificing life is acceptable. None of Kant's illustrations or examples of self-sacrifice involve someone who is the efficient cause of their death.

Indeed, if one takes seriously Kant's claim that one must not be the agent of one's death, one might argue that the death penalty is necessary because it provides a legitimate venue for the offender to sacrifice her life. Fleischacker worries that this interpretation requires some implausible 'double self' (1988: 439–40). But it does not; it requires only that one self has multiple aspects. These multiple aspects are visible in ordinary experience: they lead to the phenomena of criminals turning themselves in; even more prosaically, they lead to feelings of shame. Put differently, the distinction between, and unity of, *homo phenomenon* and *homo noumenon* can be asserted only in the practical realm. While theoretical reason will always view this relation as implausible, its judgments have no standing, in so far as they are theoretical judgments about what escapes experience. And to make such judgments extends theoretical reason beyond the limits set by the critical project. Kant briefly discusses this point in the *Metaphysics* (439).

On a related note, one might think there is a tension between my interpretation and Kant's claim that it is impossible 'to will to be punished' (*MM* 6: 335). But these words should not be taken out of context. They come in the middle of a response to one of Cesare Beccaria's abolitionist arguments. Beccaria argues that parties to the social contract could not consent to their own execution, and that capital punishment is therefore illegitimate (1963: 45). In the passage under discussion, Kant is attacking one of the basic premises of the argument: the claim that the legitimacy of punishment is based on the promises made by the parties to the social contract. Kant writes: 'if the authorization to punish had to be based on the offender's *promise*, on his *willing* to let himself be punished, it would also have to be left to him to find himself punishable and the criminal would be his own judge' (*MM* 6: 335). Kant thinks the absurdity of this view is plain, but this criticism in no way precludes the idea that criminals will their own punishment, in the sense that their rationality subjects them to the criminal law (*MM* 6: 335).

- ²⁷ I discuss these issues more fully in ‘Rule of Law Abolitionism’ (Yost 2007). Of course, one could argue that, on account of human fallibility, the death penalty can never be administered justly, and must therefore be abolished. But this strategy seems to prove too much. Fallibility infects all manners of criminal sentencing, and Kant would clearly not condone the abolition of punishment as such. While I think pressing this line of argument might be fruitful in a broader context, such a discussion is beyond the scope of this paper.

References

Parenthetical citations refer to the pagination (volume: page) of the standard German edition of Kant’s work, *Kants Gesammelte Schriften*, edited by the German Academy of Sciences (Berlin: Walter de Gruyter, 1900–). Unless otherwise noted, English translations come from the *Cambridge Edition of the Works of Immanuel Kant*. I employ the following abbreviations: *Anthropology from a Pragmatic Point of View* (A); Collins’s notes on Kant’s moral philosophy lectures (Col); *Critique of Practical Reason* (CPrR); *Metaphysics of Morals* (MM); *Religion within the Boundaries of Mere Reason* (Rel); ‘On the Common Saying: That May Be Correct in Theory, but It Is of No Use in Practice’ (TP); Vigilantius’ notes on Kant’s metaphysics of morals lectures (Vig).

- Baldus, D., C. Pulaiski and G. Woodworth. (1983) ‘Comparative review of death sentences: an empirical study of the Georgia experience’, *Journal of Criminal Law and Criminology*, 74: 661–753.
- Beccaria, C. (1963) *On Crimes and Punishments* (New York: Macmillan).
- Byrd, B. S. (1989) ‘Kant’s theory of punishment: deterrence in its threat, retribution in its execution’, *Law and Philosophy*, 8(2): 151–200.
- Clark, M. (2004) ‘A non-retributive Kantian approach to punishment’, *Ratio*, 17(1): 12–27.
- Finkelstein, C. (2002) ‘Death and retribution’, *Criminal Justice Ethics*, 21(2): 12–21.
- Fleischacker, S. (1988) ‘Kant’s theory of punishment’, *Kant Studien*, 79: 434–49.
- Hill, T. E. (1997) ‘Kant on punishment: a coherent mix of deterrence and retribution?’, *Jahrbuch fuer Recht und Ethik*, 5: 291–314.
- Hill, T. E. (2003) ‘Treating criminals as ends in themselves’, *Jahrbuch fuer Recht und Ethik*, 11: 17–36.
- Hoffe, O. (2002) *Categorical Principles of Law: A Counterpoint to Modernity* (University Park, PA: The Pennsylvania State University Press).
- Holtman, S. W. (1997) ‘Toward social reform: Kant’s penal theory reinterpreted’, *Utilitas*, 9: 3–2.
- Korsgaard, C. M. (1996) *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press).

- Merle, J. C. (2000) 'A Kantian critique of Kant's theory of punishment', *Law and Philosophy*, 19(3): 311–38.
- Murphy, J. G. (1987) 'Does Kant have a theory of punishment?', *Columbia Law Review*, 87(3): 509–32.
- Nathanson, S. (2001) *An Eye for an Eye: The Immorality of Punishing by Death* (Lanham, MD: Rowman & Littlefield).
- Potter, N. T. (2002) 'Kant and capital punishment today', *Journal of Value Inquiry*, 36(2–3): 267–282.
- Pugsley, R. (1981) 'A retributivist argument against capital punishment', *Hofstra Law Review*, 9: 1501–23.
- Radin, M. J. (1980) 'Cruel punishment and respect for persons: super due process for death', *Southern California Law Review*, 53: 1143–85.
- Sarver, V. T. (1997) 'Kant's purported social contract and the death penalty', *Journal of Value Inquiry*, 31(4): 455–72.
- Scheid, D. E. (1983) 'Kant's retributivism', *Ethics*, 93: 262–82.
- Schwarzschild, S. S. (1985) 'Kantianism on the death penalty (and related social problems)', *Archiv fuer Rechts- und Sozialphilosophie*, 71: 343–71.
- Shafer-Landau, R. (2000) 'Retributivism and desert', *Pacific Philosophical Quarterly*, 81(2): 189–214.
- Shell, S. M. (1997) 'Kant on punishment', *Kantian Review*, 1: 115–35.
- Sorell, T. (1988) *Moral Theory and Capital Punishment* (Oxford, UK: Blackwell).
- Sussman, D. (2008) 'Shame and punishment in Kant's Doctrine of Right', *Philosophical Quarterly*, 58(231): 299–317.
- Uleman, J. K. (2004) 'External freedom in Kant's *Rechtslehre*: political, metaphysical', *Philosophy and Phenomenological Research*, 68(3): 578–601.
- Yost, B. S. (2007) 'Rule of law abolitionism', *Studies in Law, Politics and Society*, 42: 53–89.