

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications - Department of
Philosophy

Philosophy, Department of

1998

"The Principle of Punishment Is a Categorical Imperative"

Nelson T. Potter Jr.

University of Nebraska - Lincoln, kjohnson6@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/philosfacpub>



Part of the [Continental Philosophy Commons](#), and the [Ethics and Political Philosophy Commons](#)

Potter, Nelson T. Jr., "'The Principle of Punishment Is a Categorical Imperative'" (1998). *Faculty Publications - Department of Philosophy*. 22.

<https://digitalcommons.unl.edu/philosfacpub/22>

This Article is brought to you for free and open access by the Philosophy, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications - Department of Philosophy by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

9

Nelson Thomas Potter Jr.

“The Principle of Punishment Is a Categorical Imperative”¹

KANT’S VIEWS

There has been a considerable renaissance in retributivism as a theory of the justification of punishment in the second half of this century. Retributivism is often defended as if it were a particularly hardy moral intuition, a basic free-standing moral principle that is underivable from any broader theory or set of principles. In this vein it is often “supported” through the presentation of outrageous and horribly cruel crimes, especially against persons, particularly murder, in order to elicit what may be thought to be the natural and appropriate emotional response, a response of anger, indignation, and desire for retribution.² Under such accounts the retributive idea has little to do with ethics thought of as a rationally defended systematic theory.³

In the history of retributivism, Kant has a prominent place. He was one of the classic defenders of a tough retributivism, at a time when the new humaneness and teleology in the theory of punishment was making its first headway with the help of (equally classic) Enlightenment writers like Beccaria. I wish to show that Kant did not regard retribution as a basic, underivable moral principle; rather, he is concerned, as far as possible, to find a rational basis for the idea of retribution, and to relate it closely to the root ideas of his moral

philosophy: the categorical imperative (CI), and the idea of respect for persons. Kant's theory of punishment has been discussed usually as a series of statements that have been considered for their *implications* for the justification of punishment; it is less often considered with respect to its basis in the broader Kantian practical philosophy.

Kant uses the phrase "a categorical imperative" usually to refer to a specific obligation that is derived from *the* categorical imperative. Thus for example, "Refrain from making any lying promises" is a categorical imperative that is established by the argument presented in the second of Kant's well-known four examples in the *Grundlegung zur Metaphysik der Sitten*.⁴ So when Kant writes that "The principle of punishment is a categorical imperative" (MS, 6:331), he is telling us that this principle (a) makes an unconditional moral demand, not one that may be altered for the sake of someone's convenience or preference, and (b) that it can be derived from some version of *the* categorical imperative. The title quotation also refers to the *principle* of punishment, which Kant identifies as the *lex talionis* (MS, 6:232). The *lex talionis* is usually identified with the formulations derived from Mosaic law, "An eye for an eye, a tooth for a tooth, a life for a life."⁵

Punishment is in Kant's view a topic in the philosophy of law or right (*Recht*). Punishment is imposed not by individuals, but by courts and judges, and it is therefore a practice that presupposes the existence of the state. Kant argues in some detail that we have a moral obligation to leave the state of nature and to enter a civil commonwealth, because it is only in a state or commonwealth that external justice, especially with respect to rights to hold property, can be assured and enforced. The state then exists as the enforcer of the rights of citizens.⁶

Kant's theory of the state and its proper functions, and of the law of property and of punishment, are all statements of moral ideals. Kant understands that actual states will at best approximate to such ideals, and at worst, will fail to exemplify some or all of these ideals.⁷ We should remember that in the discussions that follow Kant is talking about developing an ideal, and we can think about the issue of punishment in a similar vein by asking to what extent Kant's ideal is embodied or not in arrangements for handling crime and punishment in present-day societies.⁸

Because the discussion of punishment is part of the *Rechtslehre*, what we are applying is not the familiar categorical imperative as stated in the *Grundlegung*: "Act only on that maxim through which you can at the same time will that it should be a universal law" (G, 4:421). Rather we will be using the more restricted version that is appropriate for *Recht*, which Kant calls "The Universal Principle of Law (*Recht*)":

"An action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law." (MS, 6:230; the quotation marks are in Kant's text)

The kind of freedom referred to is external freedom. Inner freedom of thought and motivation is beyond the reach of *Recht*. Also, with respect to inner freedom, individuals could not come into conflict. The concern of the principle of right is the ways in which one individual's use of freedom may interfere with that of another individual.⁹

It will be noticed that the quoted principle contains the phrase, familiar from the CI, "universal law." This is the main part of the principle that establishes the connection with the original CI. The universal principle of law is a more specific version of the general CI, a version limited to external actions that might affect others. The restatement of the law in terms of coexistence of everyone's freedom is a paraphrase of the "universal law" idea. My breaking a promise will be right only if my freedom to break the promise can coexist with a similar freedom on the part of everyone else to break such promises. Since in Kant's view, such freedoms of each and every person to fail to keep promises cannot coexist, then such an action cannot be right, that is, in accord with the universal principle of right. The universal law formulations of the CI are principles of justice or fairness, and in that sense they are also principles of *equality*. Kant immediately presents an argument to show that "there is connected with Right by the principle of contradiction an authorization to coerce someone who infringes upon it" (MS, 6:231). A little later he concludes, "Right and authorization to use coercion mean one and the same thing" (MS, 6:232). This authorization amounts to an authorization to punish individuals guilty of actions that violate the law, for punishment turns out to be the only kind of coercion that Kant discusses.¹⁰

The motivation within the doctrine of "*Recht*" is not the inner motive of duty that Kant has emphasized so much in the *Grundlegung* and the second Critique. But since Kant insists that a moral precept must always be paired with a motivation adequate to make it effective in human choice (see MS, 6:218–19), we here introduce the threat of coercion and punishment to take the place of inner moral motivation, which we are not relying on in *Recht* (MS, 6:232). The laws against rape or against speeding on the interstate do not appeal to our better nature in asking us to refrain; they warn us that if we violate (and are caught and convicted) we will be punished, and the fear of punishment provides the incentive. Hence, according to Kant's way of proceeding here, in the realm of *Recht*, the concern with action is limited to the

external, that which can affect another, and hence that which can bring us into conflict with the wishes and indeed with the rights of another.¹¹

When Kant begins his main discussion of punishment (MS, 6:331ff.), he juxtaposes two ideas, that of equality and that of the law of retribution (*lex talionis*), as follows (I quote at length):

But what kind and what amount of punishment is it that public justice makes its principle and measure? None other than the principle of equality (in the position of the needle on the scale of justice), to include no more to one side than to the other. Accordingly, whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself. . . . But only the *law of retribution (lex talionis)*—it being understood, of course, that this is applied by a court (not by your private judgment)—can specify definitely the quality and the quantity of punishment; all other principles are fluctuating and unsuited for a sentence of pure and strict justice because extraneous considerations are mixed into them. (MS, 6:332)

The proper measure of the punishment, namely, equality, restores the equilibrium that existed before the crime, and that was first disturbed by the crime.

The fact that he understands the criterion for punishment in this quote in terms of the idea of equality means that Kant intends the criterion as an application of the categorical imperative. In supplementary comments added in the second edition, he writes, in a similar vein, that only the *lex talionis* "is by its form always the principle of the right to punish since it alone is the principle determining this Idea a priori." (The CI is itself *a priori*.) The phrases used to spell out the CI, "universal law," "coexistence," and so on all imply equality; the idea of equality is perhaps most explicit in Kant's general ethics in the spelling out of the idea of the kingdom of ends in the *Grundlegung*. The first formulation CI is a principle of justice or fairness, and by invoking the idea of equality in settling on the proper amount of punishment we are seeking through the punishment to restore the preexisting equality of consideration.¹²

So, Kant's insistence on the *lex talionis* is inspired by his thought that punishment itself ought to be an application of the CI. We find a standard for the appropriate quantity and quality of punishment only in the basic idea of the CI itself, the idea that we are all moral equals.¹³

Now let us look more closely at this derivation, in particular the idea that the punishment is to be quantitatively and qualitatively a sort of mirror image of the crime. One of the merits of this idea, as we've seen, is that it makes manifest to the person punished the appropriateness of a

certain quality and quantity of punishment. And yet the proposition that the punishment should mirror the crime is something other than a mere analysis of the concept of punishment. Alternative statements of moral principle here, though perhaps all incorrect, seem conceivable: that the punishment should double, or be half the quantity of the crime, or that the punishment for *any* moderately serious crime should be death, or that each criminal conviction must entail the confiscation of exactly half of one's property. The possible principles here are endless, and this is a clue that the *lex talionis* is synthetic rather than analytic. And yet there seems to be something special and intuitively plausible about the *lex talionis*, as opposed to any of the possible alternatives. In Kant's scheme of philosophical possibilities, this might suggest that the *lex talionis* is synthetic *a priori*. That in fact seems to be exactly what Kant believed.

He begins to develop this idea in a remarkable passage that develops an analogy between the realm of external freedom, and the mathematical/physical world:

The law of reciprocal coercion necessarily in accord with freedom for everyone under the principle of universal freedom is, as it were, the *construction* of that concept, that is, the presentation of it in pure intuition *a priori*, by analogy with presenting the possibility of bodies moving freely under the law of the *equality of action and reaction*. In pure mathematics we cannot derive the properties of its objects immediately from concepts but can discover them only by constructing concepts. Similarly, it is not so much the concept of Right as rather a fully reciprocal and equal coercion brought under universal law and consistent with it, that makes the presentation of that concept possible. (MS, 6:233)

This analogy between the model of the physical world, which was at the center of Kant's attention for so long in the earlier years of his philosophical development, and the political model of the reciprocal rights of persons thought of as externally related to one another in space is important. The Newtonian law that is most closely connected in this analogy with the idea of punishment is the law of the equality of action and reaction, Newton's third law, a principle that Kant thought of as a synthetic *a priori* principle.¹⁴

Hence, in Kant's view, the principle of punishment is indeed synthetic *a priori*, by analogy with the geometrical/physical construction of bodies in space, and hence is necessary and invariable, like any categorical imperative. Just as Newton's laws allow for zero tolerance in variation of effect, given the cause, so likewise appropriate punishment must fit the crime exactly, according to the *lex talionis*. But

whence in the physical model comes the practical ideal that the punishment must be *qualitatively* identical to the crime? Kant does not answer this, but I suggest that the analogy for quality is vector direction of reaction. Given the status of Newton's third law, we can understand why Kant sought no further derivation of the principle of punishment, regarding it rather as a sort of practical axiom or basic law.

Kant in the *Groundwork*, as is well known, said that violations of perfect duties are actions such that "their maxim cannot even be *conceived* as a universal law of nature without contradiction" (G, 4:424). On the other hand, the "general canon" for all duties, perfect and imperfect, is that "we must *be able to will* that a maxim of our action should become a universal law" (G, 4:424). Nevertheless, many violations of duties that we wish to call perfect duties seem to involve the failure of the broader criterion only.¹⁵ This seems to be the case with respect to Kant's clearest example of the application of the principle of law, his claim that a hereditary nobility is contrary to *Recht* (see MS, 6:329; compare "Theory and Practice," 8:291–93, 297–98), for Kant's point is that a person excluded from such an advantage could not vote for such an arrangement, that is, could not will such a law. This example is discussed further below in the section entitled "Other Examples of Application."

Now we can see more clearly the sort of "derivation" claimed for the principle of punishment. There is no contradiction in conception in a system that would have the punishment double the crime, or in a system that would punish the entire range of possible crimes with equal severity, that is, with the same punishment. After all, the *lex talionis* is synthetic *a priori*, and therefore its denial is not self-contradictory. Nevertheless the *lex talionis* is the only true application principle here, in Kant's view. Analogously, nature might, without contradiction, be such that for every action the reaction doubles it,¹⁶ or it might be such that the interior angles of a triangle are greater than 180 degrees. However, the true synthetic *a priori* proposition in each case is different, stating what we know as Newton's third law and the familiar theorem of Euclid.

So, although there is no incoherency or contradiction in having the punishment double the crime, or halve the crime, such alternative principles could not be willed by everyone. The subject being punished could not rationally will excessive punishment, and those benefiting from deterrence or those who are crime victims or friends of crime victims could not rationally will punishment that was too light. The *lex talionis*, proposing as it does *equality* of crime and punishment, is simply an extension or instantiation of the universal principle of law, which itself is already a principle of equality or fairness. With punishment, the key moral concept of equality (of initial distribution, or of

regard before the law) is simply extended to state a principle for responding to publicly wrongful actions.

Kant's Partial Retributivism

Kant is a partial retributivist only, because the function of the institution of punishment is to provide deterrence of antisocial acts.¹⁷ Punishment hence has a teleological function, a goal: Punishment serves the function of protecting the rights of all of us, by providing crime control through deterrence. The nonteleological side of punishment pertains to determining the proper quantity and quality of punishment, determining the proper distribution of punishment for those who have violated the law. This is not determined teleologically, but formally, by the principle of equality, as embodied in the *jus talionis*. This means that we may not adjust the amount of punishment to enhance deterrence, or to achieve the most efficient balance between enforcement costs and the costs of crime.¹⁸

The actual examples of the *lex talionis* that Kant provides sound to us quite harsh. Probably for that reason there was little problem within the punishment theory as Kant thought of it, of having the *lex talionis* indicating punishments that would be inadequate in their deterrent effect. More often there would be "overkill," if I may use that expression, from the point of view of deterrence; that is, the punishment would be more harsh than the needs of adequate deterrence would require. We need to consider these questions of the adequacy of the "match" between deterrence needs and the punishment dealt out by the *lex talionis*, because the criteria for each are entirely independent of each other in a partial retributivist theory like that of Kant, and it appears that they could relate to each other in ways that would cause trouble. The most likely sort of trouble I think will seldom occur, namely, that the *lex talionis* mandates a punishment that is inadequate for the purposes of deterrence.¹⁹

The Right to Punish

What is Kant's chief concern in presenting a criterion to determine the appropriate level of punishment? His concern, expressed over and over again, is that the punishment be appropriate from the point of view of the person who is receiving the punishment. Sometimes this concern is expressed in the claim that the punishee could not reasonably dispute the justice of the punishment. The phrase "the right to punish" (from MS, 6:363), as Kant uses it, also suggests concern with the punishee, for he is the individual against whom this right would be exercised. After mentioning the right to punish, Kant concludes:

To inflict *whatever* punishments *one chooses* for these crimes would be literally contrary to the concept of *punitive justice*. For the only time the criminal cannot complain that a wrong is done to him is when he brings his evil deed back upon himself, and what is done to him in accordance with penal law is what he has perpetrated on others, if not in terms of its letter at least in terms of its spirit. (MS, 6:363)

Hence, the *lex talionis* alone assures that the rights of the punishee are not violated. Again, Kant writes,

Moreover, one has never heard of anyone who was sentenced to death for murder complaining that he was dealt with too severely and therefore wronged: everyone would laugh in his face if he said this. (MS, 6:334)²⁰

In addition, there is another important qualification on the imposition of punishment, which also comes from concern for the personhood of the punishee. When such a person is executed, the punishment “must still be freed from any mistreatment that could make the humanity in the person suffering it abominable” (MS, 6:333). In this brief statement Kant opposes torture executions.

Since punishment, as our title quotation has indicated, is a categorical imperative, it should not be surprising to find it so bound up in Kant’s thinking with the idea of respect for persons. Punishment must respect the rights of the punishee. Any theory of the justification of punishment, as a part of moral theory, must have an important concern about the harm inflicted on the punishee.

Too Much/Too Little

It is easy enough to understand that punishment must not exceed the limits of what is appropriate, and that, when it does so, it would be a violation of the moral rights of the punishee. But it is more difficult to understand why punishment *short* of the maximum may not be permitted. Just how and why would it be a violation of anyone’s rights for a convicted murderer to be sentenced to a long prison sentence? Yet Kant strongly speaks out against lesser sentences, and in one of his most infamous statements, urges that death penalties against convicted murderers should be carried out even when it could serve no purpose:

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 on his punishment; for otherwise the people can be regarded as collaborators in this public violation of justice. (MS, 6:333)²¹

This sort of passage strongly suggests that Kant is a thorough retributivist, rather than a partial one, as was claimed above, for in the circumstances described, the deterrent function of punishment would seem to be nonexistent. It appears that the only reason for executing the murderers is to let justice be done (= a retributivist rationale), since there would be no continuing society that would experience the benefits of deterrence that would in other circumstances flow from such executions.

Kant himself never addresses this point, so what I suggest here is a speculative extension of what I take to be Kant's basic ideas. There are two points of justification Kant may have in mind here:

1. The crime is murder, and the scales of justice are not restored until there is an execution. Not executing might seem to make the murder *victim* something less than an equal of the murderer, since the victim is dead and the murderer is allowed to live. In other cases, when a serious crime is punished lightly, the crime victims feel badly treated, and discriminated against. Anything less than full retaliation, it might be said, denies justice and equality to the victim.
2. Joel Feinberg quoted some of these extreme statements from Kant in his "The Expressive Function of Punishment,"²² saying that such passages marked Kant as an expression theorist. The expressive function is no part of Kant's official theory, but if he did hold to it and it explains such statements, then he would be saying these things because he felt it was important to denounce, through the act of punishment, the injustice of the crime, and to reassert the innocence of the victim. However, the expressive function mostly reduces to a deterrence aspect, which is irrelevant in determining appropriate quantity and quality of punishment.

It is true that Kant thinks of the *lex talionis* as being an exact, precise standard that ideally at least allows zero tolerance. Punishment according to the proper standard of the CI should be neither too little nor too much. Such comments by Kant also may point toward the idea that the failure to carry out appropriate punishment is a violation of a

duty to *oneself*, which would reflect a lack of inner integrity or self-respect. To fail to punish to the proper degree is to fail to have an adequate hatred of criminal conduct; such an idea, which seems implicit in Kant, brings him close to being an exponent of the expression theory.

Notice that the application of the CI to derive a standard of punishment is different from the familiar four examples in the G, and other similar examples elsewhere in Kant's moral philosophy. It is different because it does not have the agent deliberating about the moral acceptability of her or his own maxim. Rather, the CI is used, not in a context of personal choice and deliberation, but in order to make an abstract moral judgment about the appropriate level (and, in Kant's view, quality) of punishment. Such a judgment would be relevant to the actions of and hence of interest to the sentencing judge, the legislator, the punishee, the crime victim, or the mere interested moral observer, in relation to their different roles.

In review, Kant's retributivism does not at all present the common idea that retribution is a basic, axiomatic, underivable moral intuition or emotion. The *lex talionis* is instead derived from the categorical imperative itself, and for this reason includes as central to it the CI idea of respect for persons, including the person to be punished. This is an abstract derivation of a moral precept, characteristic of the *Rechtslehre*, rather than a personal deliberation about one's own maxim, and hence about one's own course of action that is characteristic of Kant's ethics. Also, Kant is not a pure retributivist; the state institutions of punishment are to serve the indispensable function of crime control. In all of these ways Kant, unlike many more recent writers, "rationalizes" his retributive theory of punishment. The famous outbursts, especially about the need to execute all the condemned murderers before a society disbands, cannot be entirely understood within the scope of Kant's "rationalized retributivism," but I hope that understanding how Kant tries to justify his version of retributivism can help to make them seem less outrageous. Kant thinks that punishment cannot be less than the *lex talionis* requires, perhaps because such a punishment would reflect a lack of equal respect for the crime victim, and also a lack of adequate hatred of the wrongdoing being punished.²³

RELATED MATTERS

Now that we have presented the outline of Kant's views on punishment, emphasizing how they are derived from the categorical imperative, we will pursue some related topics.

Punishments Incompatible with Humanity

As mentioned above, Kant rejected torture executions, saying that capital punishment "must still be freed from any mistreatment that could make the humanity in the person suffering it into something abominable" (MS, 6:333). This thought, that exotic torturous forms of execution such as drawing and quartering should not be permitted, is at least one of the thoughts that moved the writers of the U.S. Bill of Rights to forbid "cruel and unusual punishments" in its Eighth Amendment.

In the early 1970s the question of whether capital punishment in and of itself may be contrary to the U.S. Constitution because it constitutes cruel and unusual punishment was the question before the U.S. Supreme Court, and in *Furman v. Georgia* (408 U.S. 238 [1972]) the Supreme Court did reject capital punishment as it then existed as unconstitutional. The five justices who formed the majority in *Furman* were divided, however. Two, Brennan and Marshall, thought that in principle capital punishment was unconstitutional. Three other justices were only addressing capital punishment laws as they then existed, and some of them were later convinced to join the majority in 1976 in *Gregg v. Georgia* (428 U.S. 153 [1976]), in upholding a new form of law that tried to guide the discretion of the sentencing authority (judge or jury) by listing mitigating and aggravating circumstances in the light of which that sentencing authority would decide between death or a lesser penalty. Our present interest, however, is in Brennan's more absolutist opposition to the death penalty as found in his *Furman* opinion. I quote at length:

Death is truly an awesome punishment. The calculated killing of a human being by the State involves, by its very nature, a denial of the executed person's humanity. The contrast with the plight of a person punished by imprisonment is evident. An individual in prison does not lose "the right to have rights." . . . A prisoner remains a member of the human family. . . . An executed person has indeed "lost the right to have rights." As one 19th century proponent of punishing criminals by death declared, "When a man is hung, there is an end of our relations with him. His execution is a way of saying, 'You are not fit for this world, take your chance elsewhere.'"

In comparison with all other punishments today, then, the deliberate extinguishment of human life by the State is uniquely degrading to human dignity.

The language of Kant and of Brennan is not precisely the same, but this conclusion seems unavoidable: the reasons that Kant gives for rejecting

torture executions (while defending capital punishment) are the same as those Brennan gives for opposing capital punishment in every case. Such treatment is incompatible with the humanity of the person being punished. Brennan and Kant disagree only concerning the scope of the state punishments that are forbidden. The only additional difference (which here makes no difference) is that whereas Kant is expressing a moral ideal, Brennan is writing as a Supreme Court justice with the authority to interpret constitutional language, which is itself usually understood as embodying certain moral ideals.

Now some may urge that the worst murderers have no humanity to preserve. Kant would have disagreed with that, it seems, since he never considered the possibility of beings with human shape but without human potential. Jean Hampton briefly considers this possibility in "The Moral Education Theory of Punishment," but does not discuss it.²⁴ Against this idea, it may be said that whenever we wish to proceed to murder or genocide in some semipublic fashion, we do so by dehumanizing those who are to die, as the Nazis did; so the hypothesis, whether correct or incorrect, presents special dangers of potential for misuse. Like Hampton, we will have to dismiss this possibility without at present considering it further.

How are we to decide who is correct, Brennan or Kant? How broad is the correct scope of the "inhumanity objection"? Does it cover only torture executions or does it cover all executions? As thus stated, it seems to me an unmanageable question that may not admit of a generally convincing unique correct answer. In any case, answering this question convincingly is also a project beyond the scope of this paper.

Suppose Brennan were right. Then Kant would have to make do with some lesser punishment such as long-term imprisonment for such "capital crimes" as murder or treason. Similar problems arise in connection with punishments Kant proposes in a second-edition appendix to the *Rechtslehre*. These additional examples are to make the point that although one may not be able to find a qualitatively similar punishment, one can find a punishment that brings back upon the criminal his own conduct "if not in terms of its letter at least in terms of its spirit" (MS, 6:363). Kant writes:

The punishment for rape and pederasty is castration (like that of a white or black eunuch in a Seraglio), that for bestiality, permanent expulsion from civil society, since the criminal has made himself unworthy of human society. (MS, 6:363)

The U.S. Supreme Court has ruled out banishment as violative of the "cruel and unusual punishment" clause (*Trop v. Dulles*, 356 U.S. 86

[1958]), and the mutilation of castration is one that many might wish to reject on similar grounds.²⁵

Lex Talionis

A related problem, in spite of Kant's suggestions for punishments "in the spirit" of the crime, is whether there are *qualitatively* appropriate punishments for all crimes. Kant has an ingenious suggestion for the appropriate punishment for theft:

But what does it mean to say, "If you steal from someone, you steal from yourself"? Whoever steals makes the property of everyone else insecure and therefore deprives himself (by the principle of retribution) of security in any possible property. He has nothing and can also acquire nothing; but he still wants to live, and this is now possible only if others provide for him. But since the state will not provide for him free of charge, he must let it have his powers for any kind of work it pleases (in convict or prison labor) and is reduced to the status of a slave for a certain time, or permanently if the state sees fit. (MS, 6:333)

If we accept this account, then we have a large class of crimes taken care of by something like imprisonment at hard labor: all property crimes, including all theft, fraud, and burglary. The contemporary American universal solution for criminal behavior of imposing long prison sentences thus gains some endorsement from Kant with respect to a large share of those now in prison.

But other crimes remain problematic. What about drug dealing? First we must decide what the gravamen of this offense is, and that would be no easy matter. What about tax evasion? Or perjury? The problem of finding appropriate mirror-image punishments is more acute in the case of crimes such as these, which may not have an individual victim. What about assault? Do we assault the offender back? What about escape from custody? The task of finding a punishment for each type of crime that is a quantitative and qualitative mirror of the crime seems to be not possible of accomplishment, especially when the demand for *qualitative* similarity is included. Even in the case of theft, where Kant urges imprisonment at hard labor, we should ask, "How long?" Will the answer depend on the amount stolen? Double time for double money? The task of constructing perfectly equivalent punishments for each crime seems beyond human capability. We could surely accomplish such a task only by loosening the ties between crime and punishment, along the lines proposed in Hyman Gross's schema

for determining appropriate punishment, where the scale of crimes is arranged and opposite it is a scale of punishments.²⁶ Crimes must be arrangeable at least ordinally, something that is difficult enough. The scale of punishments will have a certain pitch. But as Gross describes it, there is no single “verifiable” point of contact between the two scales; if there were, this would give us confidence about the correct match of other point-pairs on the two scales. A defender of *lex talionis* like Waldron does so without seeking for such a perfect, zero-tolerance match between crime and appropriate punishment.

Kant discusses a large number of examples that in his view affect appropriate punishment. He seems particularly interested in class distinctions, which he thinks will make a difference in determining appropriate punishment. These examples, which taken together are somewhat puzzling and difficult to make sense of because of their variety, are listed here to make it clear that Kant is sometimes less doctrinaire with his examples than in his more abstract pronouncements. For example, after insisting fervently that death is the only appropriate punishment for murder, he presents no fewer than three examples that are in his view justified *exceptions* to this rule! ((4), (5), and (6), below). In contrast (3) seems to be presented as a complicated argument for the death penalty for treason. (1) Kant writes of a person of high rank verbally insulting a person of lower rank. Punishment: the offender must publicly apologize, and kiss the hand of the insulted one, which will produce in him a humiliation similar to the one he produced (MS, 6:332). (2) Again, he writes of a high-status person striking a low-status person. Punishment: The offender must apologize and also must undergo solitary confinement involving hardship (MS, 6:332–33). (3) A complicated case of men of honor and scoundrels being tried together for treason (MS, 6:333–34). (4) Another complicated case where almost everyone who was a citizen was an accomplice in the murder. Solution: All should be sentenced to death, but then reprieved by executive decree, and given a lesser punishment, perhaps banishment (MS, 6:334). The cases of (5) the mother murdering her illegitimate child (MS, 6:336) and of (6) killing a fellow soldier in a duel. In both cases the killing has been motivated by a sense of honor, which would suggest a lighter penalty. Kant’s puzzling solution to both cases is that death remains objectively the appropriate penalty, though it appears to be unjust for reasons that are obscure to me (see MS, 6:336–37).

In conclusion, we can say that Kant’s abstract theory of punishment sets a very high ideal: the punishment must be perfectly matched to the crime, both quantitatively and qualitatively. There is a suggestion that, ideally at least, there should be zero tolerance, in the sense that we should be able to specify precisely what the appropriate punishment is.

Only if there is zero tolerance will the punishment be defensible to the person being punished, thereby making its appropriateness clear to her.

Punishment as Expressive and as Moral Education

This aspect of Kant's theory of punishment makes it seem similar to two theories of punishment that have been more recently proposed: the expressive theory of punishment (Feinberg) and the moral education theory (Hampton).²⁷ All three views, in somewhat different ways and for somewhat different reasons, emphasize the *symbolic* character of punishment. The expressive theory, already mentioned above, emphasizes the authoritative denunciation of the offender and his action that is accomplished through the punishment. The moral education theory emphasizes the aim of improving the moral character of the person punished through the punishment's delivering a moral message. In Kant's view the punishment is symbolic, but the primary recipient of the message is the person punished, and the message is: "This hard treatment is appropriate." Hampton is opposed to the death penalty as incompatible with the goal of moral-education,²⁸ and Kant gives little hint that he wishes or expects to produce learning in the person punished. If such an expectation were central to the institution of punishment, rather than a happy by-product in some cases, then Kant would have introduced teleological considerations into the heart of his account of punishment. If the punishment is to have a general deterrent effect, others must know of it, and probably must judge it appropriate; to this extent Feinberg (and Scheid and Byrd) is correct in judging Kant to be a deterrence theorist. But in Kant's own view deterrence is a secondary (though still essential) effect, with the primary concern being the appropriateness of the punishment in relation to the punishee.

Other Examples of Applications

Kant does not have very many examples of explicit application of the categorical imperative in the context of *Recht*. Perhaps the theory of property is another such application, at least at certain points. I will not pursue this example now, since it would involve considerable complexity. A more manageable example for our purposes is Kant's argument against the privileges of a hereditary class of nobility. Like Kant's theory of property, it also involves the key requirement of potential universal advance consent.

Kant states the requirement of advance consent most fully in "Theory and Practice" (1793, KGS 8:273–313), where he describes it as

“a mere idea of reason, one, however, that has indubitable (practical) reality” (KGS 8:297):

Specifically, it obligates every legislator to formulate his laws in such a way that they *could* have sprung from the unified will of the entire people and to regard every subject, insofar as he desires to be a citizen, as if he had joined in voting for such a will. For that is the criterion of every public law’s conformity with right. If a public law is so formulated that an entire people *could not possibly* agree to it (as, e.g., that a particular class of *subjects* has the hereditary privilege of being a *ruling class*), it is not just; however, if *only* it is *possible* that a people could agree to it, it is a duty to regard that law as just, even if the people are presently in such a position or disposition of mind that if asked it would probably withhold its consent. (KGS 8:297; also see 8:289–300; and MS, 6:328–29)

This criterion is similar to what Kant in the *Groundwork* calls “the general canon for all moral judgment of action” (G, 4:424), namely, that “we must *be able to will* that our action should become a universal law” (G, 4:424). It is *not* an instance of the more specific test used in connection with the case of the lying promise, namely, that “Some actions are so constituted that their maxim cannot even be *conceived* as a universal law of nature without contradiction” (G, 4:424). One can speculate that such *Recht*-applications present another partial picture of the idea of the Kingdom of Ends that Kant presents in the *Grundlegung*, though Kant never tells us so. Clearly, the requirement of general will-like universal consent is an alternative statement of the “universal law” conception that is found in many formulations of the categorical imperative.

This general criterion, together with the sample application to reject hereditary nobilities, represents another sample application of the Kantian principle of law, together with another description of the application process. The exercise takes place entirely in the realm of ideas, and the results of any such exercise would be conclusions about the normativity or lack thereof of actual positive laws and social institutions.

Does this general criterion work for the case of Kant’s theory of punishment? Consider the following as a sketch for a Yes answer: Who is the person whose reason is not likely to consent to punishment? Answer: the one being punished. So, in justifying punishment we must focus our attention on that one, and assure ourselves that we can justify the punishment to *him*. So the general criterion explains Kant’s approach to the justification of punishment.²⁹

Probably the main difference between such examples and the more familiar four examples in the *Grundlegung* is that the latter *ethical* applications deal with the agent's *maxims*, and hence have a background assumption about the agent's motivation that is absent in the applications of *Recht*. Moral principles of *Recht* are always mere impersonal precepts, with the motivation left unspecified, and also the direction to specific agents left unspecified. Maxims, reflecting as they do the agent's personal choices of actions, ends of action, and motives of action are absolutely central to the realm of inner freedom that Kant calls "ethics."³⁰ But judgments concerning *Recht* involve no such reference to the agent's personal choices or motives.³¹

Conclusion

The main aim of this paper has been to examine Kant's views on punishment in relation to the most central ideas of his moral philosophy, especially respect for persons, but also the ideas of the categorical imperative, and of justice and equality. Kant's views on punishment are not isolated and separable from his other ideas on ethics and politics. Arguably, this should make his views on punishment seem more interesting than they otherwise might seem, for they are shaped by the most powerful and attractive ideas that make his ethical theory as a whole a matter of so much philosophical interest. The harsh retributivism that lies on the surface of these texts is tempered by deeper-lying teleological themes: the benefits of deterrence of antisocial actions in preserving our rights, and the moral teleology of the idea of respect for persons. This discussion, also enriched with a brief consideration of Kant's argument against hereditary nobility, broadens our understanding of how the categorical imperative is to be applied outside of ethics, in the area of *Recht*.

NOTES

1. MS, 6:331. MS stands for *Metaphysik der Sitten* by Kant (1797–98). G stands for *Grundlegung zur Metaphysik der Sitten* (1785). Later references to Kant's writings are given in the text; after an abbreviation for the work, these references contain volume and page of the passage in in Prussian Academy edition of Kant's works: *Kants gesammelte Schriften, herausgegeben von der Deutschen (formerly Königlichen Preussischen) Akademie der Wissenschaften*, 29 volumes (Berlin: Walter de Gruyter [and predecessors], 1902), hereafter KGS. Most English translations of works referred to include these page numbers marginally. For translations from MS I use Mary Gregor's translation.

2. For example see Michael S. Moore, "The Moral Worth of Retribution," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987). Moore even makes use of that popular master of inflammatory rhetoric Mike Royko to make the resentment family of feelings seem to be morally attractive. It has sometimes even been urged that there may be an evolutionary basis for retribution, because a strategy of being initially open to cooperation with new persons, and then responding in a friendly fashion to friendly responses, and in unfriendly fashion to unfriendly responses (and it is here that we have the retribution-like response) may have by this time in our species-history as social animals been hardwired into us. See Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).

3. For more on this see Michael Davis's *Make the Punishment Fit the Crime* (Boulder, Colo.: Westview Press, 1992), chapter 2, "The Relative Independence of Punishment Theory."

4. What is directly established by the procedure of applying the categorical imperative is the acceptability or not of the *maxim* of action that is being tested. But if a lying promise maxim is shown to be unacceptable in a CI test then there is or seems to be a rather direct inference to the following being a categorical imperative: "Do not make lying promises."

5. See Exodus 21:23–25, where the full quote is: "And if any mischief follow, then thou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe." Cf. Numbers 35:21; Leviticus 24:17–20. For an excellent recent discussion of this idea, see Jeremy Waldron, "Lex Talionis," *Arizona Law Review* 34 (1992): 25–51.

6. This is admittedly an optimistic view of the state, or perhaps better an idealized view of the state. But Kant's view on this obligation to leave the state of nature, and on his absolute prohibition of revolution no matter how unjust the state, seems to be that any state at all is better than the state of nature.

7. For example, Kant was a republican, and an opponent of hereditary nobility. He lived in an authoritarian state that was something quite different from a republic, and was surrounded by political societies that had hereditary nobilities; in fact, further, he was as a university professor an employee of the state, and he made his peace with it, and expressed in his writings his commitment to obey its commands. For a discussion of the appropriate respective realms of obedience and scholarly freedom see the essay "What Is Enlightenment?" and also "Strife of the Faculties," both KGS 9.

8. Kant makes this point when, after offering a republican definition of the state, which, so far as it is in accord with the pure principles of Right, forms an Idea of the state: "This Idea serves as a norm (*norma*) for every actual union into a commonwealth (hence serves as a norm for its internal constitution) (MS, 6:313). As such it provides us with a standard for use in criticizing existing institutions and practices. We must strive after realizing this ideal: "it is that [ideal] condition which reason, *by a categorical imperative* [Kant's emphasis] makes it obligatory for us to strive after" (MS, 6:318).

9. For example, A's hitting and injuring B interferes with B's freedom, though it may be something that A wants to do. A's taking B's food and eating it himself again interferes with B's intended use. A complete discussion of the retributive idea would also include crimes that may not have specific victims, such as tax evasion or perjury. Although there is a mention of a "maxim" above, as in the earlier statements of the CI in the *Grundlegung*, and maxims are inner entities that state people's intentions and motives, the only concern of this principle of the *Rechtslehre* is with external action.

10. Advance prevention is of limited usefulness; burglars and rapists, for example, usually take care to avoid the presence of others who might interfere. Kant gives us a Newtonian analogy for the importance of the concept of equality:

The law of a reciprocal coercion necessarily in accord with the freedom of everyone under the principle of universal freedom, is, as it were, the *construction* of that concept, that is, the presentation of it in pure intuition a priori by analogy with presenting the possibility of bodies moving freely under the law of the *equality of action and reaction*. (MS, 6:232)

The phrase "a fully reciprocal and equal coercion" (MS, 6:232) that Kant uses in explicating this idea surely means not that everyone is coerced equally, but more like this: that everyone is subject to coercion to an equal extent, whenever he or she has performed the same sort of punishable act.

11. If wealthy A offers starving B money for food, he affects him, and his failure to offer help affects him negatively. But B has no right to A's assistance, and hence A's obligations in such cases are discussed not in the *Rechtslehre* portion of the *Metaphysik der Sitten* but in the *Tugendlehre* portion.

12. Equality of consideration may be illustrated as follows: A murders B, and is subsequently punished by being executed. The status quo ante was that A and B were equal in being alive; the status quo post is that they are equal in both being dead. The original equality is impossible to restore, since that could happen only by bringing the dead (A) back to life. But the formal equality that existed before also exists again after the punishment, and so there is a sort of restoration of equality.

13. Notice that Kant's scales of justice comparison is merely quantitative; Kant indicates he also wishes the punishment to be *qualitatively* identical as well, and in fact he defends capital punishment for murder by saying that "There is no *similarity* between life, however wretched it may be, and death, hence no likeness between the crime and the retribution, unless death is judicially carried out upon the wrongdoer" (MS, 6:333). I believe that Kant will have difficulty defending his insistence that the punishment be similar in quality as well as in quantity, but defense of this thesis will have to await another paper.

14. See Michael Friedman, *Kant and the Exact Sciences* (Cambridge, Mass.: Harvard University Press, 1992), 44.

15. This surprising result is perhaps clearest in Barbara Herman's essay, "Murder and Mayhem," reprinted as chapter 6 of *The Practice of Moral Judgment*, ed. Barbara Herman (Cambridge, Mass.: Harvard University Press, 1993).

16. I say this although any departure from the equality of action and reaction is a violation of conservation principles that are very basic to physics. If the reaction doubled the action, for example, then we would have a creation of new energy out of nothing, and the possibility of a perpetual motion machine, for instance.

17. For development of this interpretation see Don Scheid, "Kant's Retributivism," *Ethics* 93 (1983): 262–82; Sharon Byrd, "Kant's Theory of Punishment: Deterrence in its Threat, Retribution in Its Execution," *Law and Philosophy* 1 (1989): 151–200; and the author's own "Kant on Obligation and Motivation in Law and Ethics," *Jahrbuch für Recht und Ethik* 2 (1994): 95–112.

18. For example, some law that is difficult to enforce, such as vandalism, we might enforce by punishing the few we catch with excessive severity; this might be less costly than beefing up enforcement, and catching more people in the act, which might have a deterrence-enhancing effect equal to the other policy, while being more expensive. The now familiar utilitarian idea of *maximizing* efficiency of deterrence is an idea that comes into common currency only in the nineteenth century, so in Kant we have the idea of deterrence, but no idea of maximizing deterrence. In any case, such a maximizing idea is quite at odds with Kant's approach in this area.

19. Where the mandated punishment is inadequate, we might call the principle the *lex talionis*. What should count as "adequate" deterrence is a nice question. It cannot mean a punishment adequate to deter any would-be criminal, for this would mean no punishment would be adequate. How can any punishment be adequate to deter a would-be malefactor who believes, no matter how irrationally, that he will not get caught, or who is not at all thinking about possible consequences of his punishable act? Such actions seem to be "undeterrable." Very roughly, a punishment will be "adequate" in deterrent effect if it would deter most deterrable criminal acts of a given kind.

20. Even in the second Critique, where Kant only briefly introduces the subject of punishment for a paragraph, he mentions the concern about the right of the punishee:

For even though he who punishes can do so with the benevolent intention of directing his punishment to [the ultimate happiness of the one being punished], it must nevertheless be justified as punishment, i.e., as mere harm in itself, so that even the punished person, if it stopped there and he could see no glimpse of kindness behind the harshness, would yet have to admit that justice had been done and that his reward perfectly fitted his behavior. (KpV, 5:37)

21. Maybe Kant got carried away here. He sometimes gets carried away when he is writing against an opponent. (Cf. the response to Benjamin

Constant in "On a Supposed Right to Lie . . .") His juices have been brought to a boil by Marchese Beccaria (referred to at end of MS, 6:334), who opposed capital punishment, and he is tempted into an overstatement in his reply. It is tempting to say this, but this explains away rather than explaining the passage. I think personally, however, that this remains a part of the correct explanation of such extreme comments.

Kant makes a similar statement against lessening punishment earlier. It is the rest of the sentence that our title-quotation comes from:

The principle of punishment is a categorical imperative and less woe to him who crawls through the windings of eudaemonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises, in accordance with the Pharisaical saying, "It is better for *one* man to die than for an entire people to perish." (MS, 6:331–32)

So if such extreme remarks are a slip of the pen, or a temptation in the heat of a response, they are repeated slips or temptations.

22. Reprinted in Feinberg's *Doing and Deserving* (Princeton, N.J.: Princeton University Press, 1970), 95–118. Insofar as Kant has in mind the physical construction model of bodies in space in thinking of punishment, this model limits comparison with conceptions of punishment as expressive. Deterrence in the physical analogy would be like the natural consequences of an equal reaction = punishment restoring the original equilibrium of a given body with others, rather than the effects of authoritative denunciation and its psychological (e.g., deterrent) effects on others.

23. I have hinted at the difficulties, in my view, of defending, or even of working out and presenting, a version of the *lex talionis* that provides for both quantitative and qualitative likeness between crime and punishment. If the requirement of qualitative likeness is dropped, arguably the *lex talionis* becomes little more than the retributive idea that the punishment should fit (a quantitative concept) the crime. I will discuss Kant's presentation of the *lex talionis*, and the numerous examples he includes later in this chapter.

24. See her comment directly after her footnote 22. The killing of human beings when it is done after the model of the destruction of excessive numbers of dogs or cats, or badly injured horses, for example, would be something other than punishment. Such actions would be clearly incompatible with any positive claims concerning the humanity of the individual being thus killed.

25. Proposals for castration are often presented as an alternative to life in prison to be undertaken only with the consent of convict. Even so, it raises troubling issues. Another problem with the punishment for "bestiality" is that it approximates to being a victimless crime that adversely affects the freedom of no other person; it seems that it would at worst be a violation of a duty to oneself, especially given Kant's views on the moral standing of animals.

26. See his "Culpability and Desert" from *Archiv für Rechts und Sozialphilosophie* 19 (1983). This essay is reprinted in most editions of *Philosophy of*

Law, 4th ed., ed. Joel Feinberg and Hyman Gross (Belmont, Calif.: Wadsworth, 1991), 669–76, or example.

27. Feinberg, *Doing and Deserving*; Jean Hampton, “The Moral Education Theory of Punishment,” *Philosophy and Public Affairs* 13.3 (1984): 208–38. Compare Herbert Morris, “A Paternalistic Theory of Punishment,” *American Philosophical Quarterly* 18.4 (October 1981). Insofar as Kant has in mind the physical/geometrical model of objects in space in equilibrium in thinking of punishment, the comparison of Kant’s views with the moral education view will be limited. According to the geometrical model individual agents are thought of simply as Newtonian bodies; hence the idea of the “education” of such bodies is quite foreign to the analogy.

28. See the discussion around footnote 22 of Hampton, “The Moral Education Theory of Punishment.”

29. Look at Kant’s interesting discussion of issues related to this point at MS, 6:335. There Kant is replying to Beccaria’s criticism of the death penalty that provision for such could never be contained in the original contract, because individuals could never consent there to lose their own life in case they murdered someone. Kant’s reply is to make a distinction between the aspect of the person that would so consent (*homo noumenon*) and the aspect that would not (*homo phenomenon*): “The chief point of error (*proton pseudos*) in this sophistry consists in its confusing the criminal’s own judgment (which must necessarily be ascribed to his *reason*) that he has to foreit his life with a resolve on the part of his *will* to take his own life, and so in representing as united in one and the same person the judgment upon a right [*Rechtsbeurteilung*] and the realization of that right [*Rechtsvollziehung*]” (MS, 6:235). There is a further discussion of issues concerning application and consent earlier in this paper.

30. See the author’s “Maxims in Kant’s Moral Philosophy,” *Philosophia* 23.1 (July 1994): 59–90, for further discussion of these points.

31. Sidney Axinn gave me useful comments on an early draft of this essay. A shorter version of it was presented to the Pacific Division of the American Philosophical Association on March 31, 1995, with useful and provocative comments by Jacqueline Marina and from the floor. Most important, very helpful critical comments were given me on the penultimate draft of the paper by Don Scheid, and all along the way I discussed the project with my colleague Mark van Roojen.