

Five Perspectives on Holding Wrongdoers Responsible in Kant

Benjamin Vilhauer
City College and Graduate Center, City University of New York, USA
bvilhauer@ccny.cuny.edu

Abstract

The first part of this paper surveys five perspectives in Kant's philosophy on the quantity of retribution to be inflicted on wrongdoers, ordered by two dimensions of difference—whether they are theoretical or practical perspectives, and the quantity of retribution they prescribe: (1) *theoretical zero*, the perspective of theoretical philosophy; (2) *practical infinity*, the perspective of God and conscience; (3) *practical equality*, the perspective of punishment in public law; (4) *practical degrees*, the perspective we adopt in private relations to others; and (5) *practical zero*, a perspective I argue is entailed by Kant's doctrine of strict right, which is his justification of coercing compliance with public law. Kant acknowledges 1-4, but not 5. The second part draws on Kant's account of the burden of proof in criminal law to argue that Kant is wrong to adopt 3 in responding to criminals, and that we ought to adopt 5 instead.

Introduction

This paper has two main parts. First, it offers a novel but textually grounded distinction between five perspectives in Kant's philosophy on holding wrongdoers responsible, or, more specifically, five perspectives on prescribing retribution. Second, it argues that Kant is mistaken about which perspective is appropriate for responding to criminals in public law. In the first part, the perspectives are labeled, ordered, and explained in terms of two dimensions of difference—whether they are *theoretical* or *practical* perspectives, and the *quantity* of retribution they prescribe, which ranges from zero to infinity: (1) *theoretical zero*, the perspective of theoretical

reasoning; (2) *practical infinity*, the perspective of God and conscience; (3) *practical equality*, the perspective of punishment in public law, which requires us to return like for like; (4) *practical degrees*, the perspective of private relations to others, which factors in sensible incentives to act wrongly and prescribes less-than-equal retribution; and (5) *practical zero*, the perspective of Kantian “strict right”, which cannot regard subjects of coercion as deserving retribution because it cannot regard them as violating obligations. Perspectives 1-4 are features of Kant’s theory of retribution which he explicitly acknowledges, though sometimes briefly. Kant does not acknowledge 5 (practical zero), but he is nonetheless committed to it. Kant’s implicit view is that perspectives 1-4 combine into a consistent account of retribution despite the different quantities of retribution they prescribe, because they apply to different domains. This paper argues that combining 1-3 and 5 yields a consistent account with a stronger moral justification.

The second part of the paper addresses criminal law’s burden of proof for seriously harmful retribution, which Kant describes as the “greatest possible^t moral and logical certainty” (VE 27:566). (Kant himself rarely uses the expression “criminal law”, but in this paper it will often be used to refer to the part of public law that responds to criminals.) It argues that there are strong reasons to prefer the less-stringent practical perspectives to the more-stringent practical perspectives in criminal law, and thus strong reasons to adopt the least-stringent practical perspective (practical zero).

1. Five Perspectives on the Quantity of Retribution

1.1. The theoretical zero perspective: referring imputation to the empirical character

The theoretical zero perspective is part of Kant's theoretical philosophy. Transcendental idealism holds that appearances must be synthesized in space and time according to the categories of the understanding in order for experience of an objective world to be possible. The Second Analogy argues that events must be governed by deterministic causal laws for our experience of objective temporal order to be possible. Kant thinks this also holds for the mental events which make up our phenomenal deliberations and volitions. He holds that if this deterministic causal account were the *complete* account, then we would not have free will or moral responsibility, we would not deserve anything based on our actions, and retribution would be unjustified. The theoretical perspective on the world cannot accommodate anything beyond the deterministic causal account.

Kant's term for holding agents morally responsible for actions is *imputation*, and he holds that from the theoretical perspective, "[our] imputations can be referred only to the empirical character" (A551/B579n).¹ The empirical character is the phenomenon of a human being's agency, "a part of the world of sense" and thus subject to "causal connection, in accordance with

¹ References to Kant's texts are given by volume and page number of the Academy Edition. Translations are from the Cambridge Edition of the Works of Immanuel Kant. "“" in passages from Kant indicates alteration of the Cambridge translation to reflect my own translation. Abbreviations are as follows. **A/B**: *Critique of Pure Reason*. **AP**: *Anthropology from a Pragmatic Point of View*. **2C**: *Critique of Practical Reason*. **3C**: *Critique of the Power of Judgment*. **CE**: Collins notes from Kant's Ethics lectures. **HM**: Notes from Kant's Metaphysics lectures by Johann Gottfried von Herder. **MM**: *The Metaphysics of Morals*. **NF**: *Notes and Fragments*. **RR**: *Religion within the Boundaries of Mere Reason*. **TP**: *On the Common Saying: That May Be Correct in Theory, But It Is of No Use in Practice*. **VE**: Notes from Kant's Ethics lectures by Johann Friedrich Vigilantius. **VM**: Notes from Kant's Metaphysics lectures by Johann Friedrich Vigilantius.

all the laws of determination”, such that “all its actions would have to admit of explanation in accordance with natural laws” (A540/B568). Thus “in regard to this empirical character there is no freedom” (A550/B578). So referring imputation to the empirical character can only be (as it were) a dissolution of imputation into merely causal responsibility. Kant shows how this works with an example about a “malicious lie”:

one goes into the sources of the person's empirical character, seeking them in a bad upbringing, bad company, and also finding them in the wickedness [*Bösartigkeit*] of a natural temper insensitive to shame, partly in carelessness and thoughtlessness...one proceeds as with any investigation in the series of determining causes for a given natural effect. (A554/B582)

Thus even from the theoretical perspective, we can represent the motivational states which lead to wrongful actions as involving *Bösartigkeit*, which Kant also uses to refer to wickedness from the practical perspectives described below. But from the theoretical perspective, we must represent it (and the agents it affects) as a completely deterministic phenomenon, so that it serves as a ground for explanation and prediction but not a ground for deserving suffering. Kant understands this perspective as *mechanistic*. In the second *Critique*, he writes that “all necessity of events in time in accordance with the natural law of causality can be called the mechanism of nature”, though his empirical dualism entails that “it is not meant by this that the things which are subject to it must be really material *machines*”—a deterministic will would be an *automaton spirituale* rather than *materiale* (2C 5:95). Mechanism rules out imputation (2C 5:95, 98), and by extension, retribution.

However, transcendental idealism also implies a distinction between phenomena and noumena, and this implies that our agency may run deeper than the empirical character: we may have what Kant calls *transcendental freedom* by virtue of a non-deterministic “intelligible character” which is the ontological substrate of the empirical character (or, if one adopts a “two

standpoints” interpretation, an equally metaphysically significant correlate of the empirical character such that the correlation is not a matter of ontological grounding). Kant thinks that, despite the determinism of the empirical character, transcendental freedom would make us morally responsible and deserving, thus justifying imputation. But an intelligible character could never be an object of theoretical knowledge, so its possibility from the theoretical perspective gives us *no reason to believe it is actual*, and cannot theoretically justify retribution.

Kant later claims knowledge of transcendental freedom, but on the basis of practical rather than theoretical reason. Practical reason gives us knowledge of laws about what we *ought* to do which are different from the causal laws we know theoretically. The epistemology of transcendental freedom which Kant holds from the second *Critique* onward argues from knowledge that we are bound by moral “oughts”, and the “ought” implies “can” principle, to the conclusion that we know we *can* act as we ought (2C 5:31). He claims this yields knowledge that we are transcendently free (2C 5:30).

There are reasons to doubt this epistemology, since there is no basis in experience for it. Kant acknowledges that the purported binding of the moral law has “exactly the same inward effect, that of an impulse to activity”, as a “sensible impulse” (2C 5:116). But this paper endeavors to work within this epistemology. This epistemology allows that we have a smattering of theoretical knowledge about what transcendental freedom *would* be like *if* it were actual, but holds that most of our knowledge is practical. We know theoretically that transcendental freedom would not have a determinate place in the temporal order, and would relate to empirical phenomena of actions. But we can only know its actuality, the phenomena within its scope (e.g. moral action but not bone growth), and its power (the degree of willpower or self-control it gives us) by knowing the practical laws governing our practices. Thus we can

only know we have the transcendental freedom necessary to justify any particular quantity of retribution by knowing the laws governing our practices of retribution. The laws say different things from the four practical perspectives. Kant thinks perspectives 1-4 combine into a consistent account despite the different quantities of retribution they prescribe, because they apply to different domains. According to this paper, combining 1-3 and 5 yields a consistent account which is better morally justified.

1.2. The practical infinity perspective: God and conscience

According to *practical infinity*, wrongful actions make us infinitely culpable and deserving of infinite punishment. Kant sets out this perspective in the *Religion*, but it has received little scholarly discussion.² There, Kant argues that “transgression of the moral law” involves “evil [*Böses*]” in our “*disposition*” and “maxims in general (in the manner of *universal principles* as contrasted with individual transgressions)”, and thus “brings with it an *infinity* of violations of the law, and hence an *infinity* of guilt [*Schuld*]” for which we must expect “*infinite* punishment [*Strafe*]” (RR 6:72).

This argument depends on Kant’s conception of maxims. Kant often claims that whenever we make a free decision to act, we explicitly or implicitly act on a freely adopted principle which encapsulates our reasons for acting, a principle which Kant calls a “maxim”. Henry Allison calls this the *incorporation thesis* (See RR 6:23-4, and Allison, *Kant’s Conception*, 276). Kant’s commitment to this is not always clear. In discussions of frailty (RR 6:29) and crimes committed “as an exception to the rule” (MM 6:321n) he suggests we can freely make exceptions to maxims. He appears to bracket this idea when discussing infinite

² See Palmquist (*Comprehensive Commentary*, 322) for a different interpretation.

guilt. While there is debate about what maxims are, it is clear that they are general. A maxim adoption is not a resolution to take a particular action in a particular circumstance, but instead to take a kind of action in a kind of circumstance. An intuition supporting the incorporation thesis is that a purported principle of action limited to a particular action in a particular circumstance which did not extend to other possible actions and circumstances would not be a *reason* to act, but instead something like a whim. Practical infinity is based on the idea that a wrongful action is based on a wrongful maxim, whose generality implies a resolution to act wrongly in an infinite number of possible circumstances. Thus any wrong entails desert of infinite retribution.

Kant thinks practical infinity is only appropriate for God's perspective on us, and our perspectives on ourselves. In the passage just discussed, Kant specifies that infinite punishment may be "exclusion from the Kingdom of God", or may be meted out in a *conversion* which is an activity of conscience (perhaps facilitated by God, since Kant thinks we must represent the inner judge of conscience as God [MM 6:439n, 440]). Conversion purges us of *radical evil*, the disposition to prioritize self-love over duty; through conversion, we reverse this prioritization and resolve to satisfy self-love only when morality permits it. Kant argues that infinite punishment "must be thought as adequately executed in the situation of conversion itself...[t]he emergence from the corrupted disposition into the good is in itself already sacrifice (as... 'the crucifying of the flesh')" (RR 6:73-4). While God and conscience should punish infinitely due to the generality of evil maxims, Kant writes that "it is otherwise before a human court, which takes only the individual crime into account, hence only the act and anything related to it, not the universal disposition" (RR 6:72). Kant does not explicitly state that practical infinity is not appropriate for extrajudicial retributive interactions between human beings. It is, however, implicit in his texts that the most severe retribution that can be justified in human interactions is

regulated by the court (see below), so if practical infinity is inappropriate for the court, it is also inappropriate for extrajudicial human interactions.

1.3. The practical equality perspective: punishment according to public law

Practical equality is guided by the principle that crimes ought to be punished with a severity equal to that of the crime. Kant calls this the “principle of equality” or the “*law of retribution*”, according to which the “position of the needle on the scale of justice” must “incline no more to one side than to the other”: “whatever undeserved evil you inflict upon another...you inflict upon yourself...If you insult him, you insult yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself” (MM 6:332). Kant thinks this principle of equality determines “what kind and what amount of punishment...public justice makes its principle and measure”, and “is applied by a court (not by your private judgment)” (ibid.).

Kant offers four distinguishable reasons why the public court must adopt practical equality. First, he suggests that the principle of equality is necessary to avoid treating criminals as mere means:

Punishment by a court...can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society...For a human being can never be treated merely as a means...[W]oe to him who crawls through the windings of eudaimonism in order to discover something that releases the criminal from punishment or even reduces its amount by the advantage it promises[.] (MM 6:332-3)

Kant’s point is that a consequentialism which determines the amount of punishment in terms of its utility inevitably treats criminals as mere means. In this passage, he seems to imply that the principle of equality is the only alternative.

Kant's second reason is that only the principle of equality ensures that "each has done to him what his deeds deserve... in proportion to his *inner wickedness* [*Bösartigkeit*]" (MM 6:333)." From the practical equality perspective, we refer to *Bösartigkeit* not as a ground for explanation and prediction, as we do from the theoretical zero perspective, but instead as a ground for deserved harm—though in contrast to practical infinity, as a ground for finite harm.

Kant's third reason is that prescribing less-than-equal punishments in some cases would violate universality in the rule of law. According to the Vigilantius Ethics, "[n]o remission is to be thought of... where a universal law is the guideline, whose suspension on behalf of any individual would establish a general claim to the same effect" (VE 27:553).

Kant's fourth reason is that public lawgiving must be determined "with mathematical exactitude" (MM 6:233), and

only *the law of retribution*... can specify definitely the quality and the quantity of punishment; all other principles are fluctuating... because extraneous considerations are mixed into them. (MM 6:332)

Exactitude is fundamental to Kant's distinction between right and ethics, which he sees as different legislations of practical reason. Public law must be coercively enforced on others, and is an expression of right, whose duties must be determined with precision. Ethics legislates duties of virtue, which are enforced by each of us upon ourselves, and sometimes grant latitude (MM 6:233, also see 6:390-4). An intuition supporting Kant's idea about exactitude in public law is that we cannot reasonably be expected to acknowledge others' authority to do things to us against our will without knowing the exact scope of that authority. There are, however, puzzles about how retribution can play a role in Kant's philosophy of right, which are addressed below.

1.4. The practical degrees perspective: private relations to others

The *practical degrees* perspective derives from Kant's doctrine of degrees of imputation, which has received little scholarly discussion until recently.³ Kant addresses this doctrine in publication only briefly, in the *Metaphysics of Morals*, but unpublished texts provide more detail. The Collins and Vigilantius ethics lecture notes introduce the doctrine with the same words: “[d]egrees of imputation depend on the degree of freedom” (CE 27:291, VE 27:567). The *Metaphysics of Morals* explains that “*Subjectively*, the degree to which an action *can be imputed* ...has to be assessed by the magnitude of the obstacles that had to be overcome” (MM 6:228). Sensible incentives can make it difficult to act morally. This difficulty increases our merit if we act morally, and diminishes our culpability if we act wrongly (MM 6:228, CE 27:291, VE 27:567).

Kant's point about the *subjectivity* at issue in degrees of imputation is not explained in the *Metaphysics of Morals*, but lectures and notes provide help.⁴ Sensible obstacles do not change the degree to which we are obligated—our obligations are objective matters of practical rationality, and the internal relations between the moral law and freedom make them objective conditions of freedom in a crucial sense. But there are also “subjective conditions of freedom”, including “the ability to act”, and in “the absence of these subjective grounds there is no imputation” (CE 27:291). In a note, Kant explains that

[t]he moral precepts are valid for every rational and free being, let their inclinations be what they will...The obligation is also the same for all degrees of inclination to the contrary; only the imputation is different, for the latter concerns

³ See Blöser (“Degrees of Responsibility”) and Fabbianelli (“Kant’s Concept”) for helpful discussions.

⁴ This paper’s account of this subjectivity largely coincides with Blöser (“Degrees of Responsibility”), though it draws on some different texts.

to what extent the action can be attributed to the subject himself, i.e., to his freedom. (NF 19:135, *Reflexion* 6698)

Kant's doctrine of degrees of imputation has important implications for retribution. In the *Metaphysics of Morals* he writes that "[t]he rightful effect of what is culpable is punishment" (MM 6:227), and three paragraphs later argues that culpability comes in degrees. This implies that rightful punishment comes in degrees, indexed to degrees of culpability.

What is the range of degrees of imputation? The Herder metaphysics notes set its lower boundary as the "vanishingly small":

Whatever applies in general to a certain magnitude that can become smaller, this also applies to it if it is vanishingly small.—All free actions are imputable—consequently the smallest degree of this, the natural actions, are also imputable. (HM 28:41)

It is reasonable to assume that "natural actions" refers to instinctive actions over which we have some control, but little—we have little control over breathing, but one can imagine trying to hold one's breath until one faints if it were necessary to save someone's life. Thus this passage suggests that any action within our control can be imputed to some degree, but the degree is sometimes "vanishingly small".

Kant does not state the upper boundary of degrees of imputation. The earlier discussion of practical infinity might suggest that the upper boundary is infinity. But Kant only draws on the degrees perspective to prescribe less-than-equal retribution when we judge and punish other people privately, that is, in ways not regulated by public law.⁵ In the Collins Ethics discussion of

⁵ Fabbianelli diverges on this point: he writes that Kant thinks the degrees perspective is "only valid for...empirical-psychological reflection" ("Kant's Concept", 209). Blöser ("Degrees of Responsibility") thinks Kant intends the degrees perspective to regulate some actual practices.

degrees of imputation, Kant discusses “frailty”⁶ and “infirmity”, or “[f]ragilitas and infirmitas humana”, as a “ground...for diminishing imputation”, and says that

Fragilitas humana can...never be a ground, *coram foro humano interno* [before the inner human tribunal], for diminishing imputation. The inner tribunal is ...without regard to human frailty[.]...*Fragilitas* and *infirmitas humana* can only be taken into account when judging the actions of other people...Man, as a pragmatic lawgiver and judge, must take *fragilitas* and *infirmitas humana* into consideration when dealing with others, and remember that they are only human[.]...*Imputatio valida* is a legally effective imputation, whereby the *effectus a lege determinato* [effect determined by law] is set in train by the *judicium imputans* [sentencing authority]. We can pass judgement on all men, and anyone may do so, but we cannot sentence them, since our *imputatio* is not *valida*, which is to say that my judgement does not have the authority to set in train the consequences *a lege determinata*. (CE 27:295-6)

By the “inner tribunal”, Kant means conscience, which he describes as an inner court where the judge is represented as both God and an aspect of oneself (MM 6:439n, 440). If Kant is implicitly aggregating all grounds for diminishing imputation into frailty and infirmity, then he is claiming that judgments of conscience (and presumably God’s judgments) should not use the degrees perspective to diminish imputation. Instead, humans should do this when judging others. His point that anyone may pass judgment on all men, but we may not set in train the consequences determined by law, suggests both that the degrees perspective is not relevant for public law, and also that retribution in extrajudicial relations between humans must be less severe than retribution in public law. This is supported by a passage later in the Collins Ethics, where Kant suggests that we are not only permitted, but also obligated, to judge and punish others extrajudicially with punishments less severe than those of public law:

Men are designed for the purpose of passing judgement on others...for otherwise, in matters outside the scope of external legal authority, we might not stand at the bar of public opinion as we do before a court of law. If somebody, for example,

⁶ Kant may not mean the same thing by “frailty” here as he does in the *Religion*: CE 27:293 says that “frailty consists, not only in its want of moral goodness, but also in the prevalence therein of even the strongest principles and motivations to ill-doing”.

has brought shame upon a person, authority does not punish it, but others judge, and also punish him, although only insofar as it lies in their power to do so, and hence no violence is done to him. People ostracize him, for example, and that is punishment enough. (CE 27:450)

Kant appears to hold that public lawgivers may also adopt the degrees perspective, but only when crimes affect just their private interests:

We say of the human law-giver, when soliciting pardon from him, that he should *temper justice with mercy*[.]...No remission is to be thought of...where a universal law is the guideline, whose suspension on behalf of any individual would establish a general claim to the same effect. But where...a statutory law is in question, depending solely on the whim of the lawgiver, and in which only...his private interest[s]...and not those of the state, are at issue, then in that case the law-giver, by his own will, may relax the obligation that he imposes. (VE 27:553)

An example of the statutory laws Kant has in mind might be lese-majesty laws prohibiting giving offense to a ruler. Here, the ruler would be permitted to punish violations to a lesser degree than the laws prescribe. These passages make it clear that Kant thinks we ought to adopt the practical degrees perspective toward others, but only in the context of private relations.

1.5. The practical zero perspective: coercing compliance with public law

The practical zero perspective is implicit in Kant's doctrine of strict right, which is a key concept in Kant's *Doctrine of Right* (MM 6:229-372). Strict right can be introduced via some other key concepts in Kant's philosophy of right. Kant thinks we have an "innate right" to "freedom (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with a universal law", which includes "innate equality, that is, independence from being bound by others to more than one can in turn bind them" (MM 6:237). This foundational right ramifies into more specific rights, including rights to property and safety from violence. Kant thinks the vulnerability of rights in the state of nature

implies not only permission, but also an obligation, to leave the state of nature: Kant's "postulate of public right" says that "when you cannot avoid living side by side with all others, you ought to leave the state of nature and proceed with them into a rightful condition" (MM 6:307). A rightful condition is structured by what Kant calls the "Universal Principle of Right": "Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law" (MM 6:230). If my action is right, then "whoever hinders me in it does me *wrong*; for this hindrance (resistance) cannot coexist with freedom in accordance with a universal law" (MM 6:230-31). Coercion which prevents the hindrance of rightful acts is itself rightful: "if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a *hindering of a hindrance to freedom*)" is right (MM 6:231). Universal reciprocity demands that *everyone* play a role in coercing *everyone else* to act in ways that do not hinder freedom, though completing the establishment of the rightful condition requires a civil constitution in which we transfer our individual duties to coerce to ruling authorities.

The practical zero perspective derives from the way we regard people as subjects of coercion. As mentioned earlier, Kant thinks there is a distinction between the legislations of reason in ethical lawgiving, the topic of the *Doctrine of Virtue*, and juridical lawgiving, which he also calls *external* lawgiving, and is the topic of the *Doctrine of Right*. The duty to coerce requires us to constrain each other's external actions, but does not permit us to try to force others to follow the law because it is moral. This is central to Kant's anti-paternalism and his understanding of the scope of legitimate coercion. It is in terms of this idea that Kant introduces his concept of *strict right*:

[S]trict right [is]... based on everyone's consciousness of obligation in accordance with a law; but ... this consciousness may not and cannot be appealed to as an

incentive[.]...Thus when it is said that a creditor has a right to require his debtor to pay his debt, this does not mean that he can remind the debtor that his reason itself puts him under obligation...it means, instead, that coercion which constrains everyone to pay his debts can coexist with the freedom of everyone, including that of debtors, in accordance with a universal external law. (MM 6:232)

Kant's emphasis on acting for the sake of duty elsewhere in the practical philosophy may make it puzzling that we are not permitted to try to make people follow law because they are obligated to follow it. It is intuitive to think that we are not *obligated* to appeal to others' obligations from the perspective of right. If someone is hindering my rightful freedom of action, I have a right to hinder him from doing so even if he does not feel the force of obligation. But it does not follow from this that I am not *permitted* to appeal to his obligation. The reason I am not permitted derives from a surprising view about obligation in strict right:

[R]ight need not be conceived as made up of two elements, namely an obligation in accordance with a law and an authorization of him who by his choice puts another under obligation to coerce him to fulfill it. Instead, one can locate the concept of right directly in the possibility of connecting universal reciprocal coercion with the freedom of everyone...That is to say, just as right generally has as its object only what is external in actions, so strict right, namely that which is not mingled with anything ethical, requires only external grounds for determining choice[.] (MM 6:232)

Here Kant distinguishes (i) obligations of the coerced from (ii) the authority of the coercers, and tells us that the concept of strict right does not include (i). In other words, strict right does not contain obligations on the part of the coerced. So the reason we are not permitted to appeal to obligations of the coerced in strict right is that strict right contains no such obligations. How can this be? This passage explains that strict right is a domain which gives us a duty to coerce others to comply with law using *external* grounds for *determining* their choices. Such determination, when successful, gives others *no alternative* to following the law. Kant thinks it can only be true that I *ought* to do x if (i) I *can* do x, *and* (ii) I am able to *not* do x. His best-known expressions of (ii) include G 4:449 and MM 6:222, but his clearest is at VM

29:1016: “every imperative is *subjectively contingent*, i.e., it is connected in its determination with the *possibility for the subject to deviate from the rule and to do the opposite*.” The duty to externally determine others to follow law is a duty to make such deviation impossible, so we cannot represent the coerced as obligated.⁷ Kant’s view is *not* that we have no obligation to follow law—we clearly have *ethical* obligations to follow law (MM 6:232). But ethical obligations are irrelevant in strict right. Since we all have an obligation to establish a rightful condition, strict right contains obligations, but they are obligations we have as coercers, not as subjects of coercion.

Kant explains that coercive determination in public law operates through *pathological* rather than *ethical* incentives:

All lawgiving can ... be distinguished with respect to the incentive...[L]awgiving which makes an action a duty and also makes this duty the incentive is *ethical*... [L]awgiving which does not include the incentive of duty in the law...is *juridical*...[and its] incentive...must be drawn from *pathological* determining grounds of choice, inclinations and aversions, and among these, from aversions; for it is a lawgiving, which constrains, not an allurements, which invites. (MM 6:218-9)

The Vigilantius Ethics offers further detail about pathological incentives, in an example about a wine-dealer who “finds it expedient” to mix sweet but poisonous sugar of lead into the wine he sells (VE 27:522). Such an agent is “acting on a principle that is a motive of his subjective inclination”, and must “be compelled pathologically”: a “constraint will be needed to counterbalance his maxim of selfishness, and destroy his motive for adulterating the wine. We put him in fear of the strictest controls, and of punishment” (VE 27:522). This discussion makes

⁷ This is closely related to what Marcus Willaschek calls the *non-prescriptive* character of public law (“Which Imperatives”), though his interpretation differs in some respects from the one presented here.

the point that the coerced are not obligated even more directly: “when another is to be compelled pathologically”, he can “be necessitated, but never obligated” (VE 27:521).

Kant’s claims about determining others’ choices through sensible incentives can be puzzling in light of what can seem to be a central idea in Kant’s theory of freedom—that our choices *cannot* be determined by sensible incentives. Kant often calls this a “negative” property of freedom (e.g. G 4:446, 2C 5:33, MM 6:226). But Kant’s point in these remarks is that we cannot be determined by external grounds to *violate* law, because we always retain the autonomy to act morally. Rightful coercion applies sensible incentives that motivate *compliance* with law and outweigh sensible incentives that motivate violation. When we recognize that public laws are rightful, and the balance of sensible incentives favors compliance, we can only act wrongly if we can “incorporate evil qua evil for incentive into [our] maxim” (RR 6:37), so that “resistance to the law” is “elevated to incentive” (RR 6:35). But Kant thinks only “diabolical” beings can do this, and thinks humans are not diabolical (RR 6:35-7).

As things stand in contemporary society, we often fail to create sensible incentives sufficient to coerce compliance with the law. Since some people *do* commit crimes, they clearly *can*, and they *ought not* have done so from the perspective of ethics. It may therefore seem that strict right must also have conceptual space to accommodate obligations on the part of people who are insufficiently coerced to follow the law. However, strict right appears to be regulated throughout by an ideal of sufficient coercion. Kant never suggests that strict right allows obligations on the part of the coerced to emerge when coercive pressure is insufficient. One way to make this intuitive is to focus on the obligation which clearly *is* a part of strict right, that is, coercers’ obligation to prevent crime. Strict right shows us that we are all sufficiently externally determinable that we can prevent each other from committing crimes with the right sensible

incentives, and that we are obligated to do so. Coercers who violate this obligation and allow crimes to occur, and then try to shift responsibility by arguing that criminals have also violated obligations, are irresponsibly seeking to mitigate their own responsibility for crimes. Consider a chief of police who blames a murder he could have prevented on the murderer's violation of his obligation not to murder. If we consider obligations of ethics along with obligations of right, his complaint is not without basis. But it is irrelevant to the perspective he ought to take on crime given his role, and it is an effort to duck his responsibility. This point may not seem to generalize to all crime—it may seem that limitations in our technology and social practices mean that we *cannot* stop all crime and thus cannot be *obligated* to do so. But to the best of our knowledge, these are contingent and temporary limitations, and they play no role in the ideal that regulates strict right and the obligations it gives coercers. The ideal of sufficient coercion that regulates strict right draws on our in-principle capacity to stop all crime, a capacity we learn we have from the simple fact that we are all externally determinable enough to be caused to follow the law by the right sensible incentives. We would need considerable technological and social progress to fully exercise this capacity, but Kant recognizes that the fact of the capacity has important implications for criminal justice, though he sees these implications differently than the arguments below claim he should.

The way strict right prescind from obligations on the part of the coerced gives us a short argument that strict right entails a practical zero perspective on retribution. Retribution entails holding people morally responsible for wrongs. But we can only hold people morally responsible for wrongs when we can appeal to obligations they have violated. We cannot do so from the perspective of strict right, so we cannot justify retribution from the perspective of strict right. Kant himself does not draw this conclusion, but it is implicit in strict right.

There is a longer argument for a practical zero perspective in strict right which Kant himself partly endorses. Strict right requires us to necessitate each other to comply with law through pathological incentives, but a power of choice which can be pathologically necessitated is an *animal* power of choice:

a power of choice is sensible insofar as it is pathologically affected (through moving-causes of sensibility); it is called an animal power of choice (*arbitrium brutum*) if it can be pathologically necessitated. (A534/B562, also see A802/B830).

Because of this, Kant calls animals *automata spiritualia*:

when a dog ravages some carrion, movement begins in him which is not caused by the odor in accordance with mechanical laws but through the arousal of desire. In animals, however, this is just as much of an external necessitation as it is in machines; thus they are called *automata spiritualia*. (NF 17:313-4, *Reflexion* 3855)

Kant's point that the dog's movement is not caused according to mechanical laws, but is nonetheless mechanistic, is that while animals are not *material* machines (3C 5:464n), their actions are necessitated by laws of sensibility which are *deterministic* and *external* in the same sense as laws governing material machines. In the second *Critique*, Kant argues that we too would be *automata spiritualia* if our willings were necessitated by determining grounds that precede them in time and are no longer under our control when we act, and argues that this would imply that we lack the "freedom which must be put at the basis of all moral laws and the imputation appropriate to them" (2C 5:96-7, also see VE 27:501-2). The duty of strict right is a duty to determine others through such grounds.

These connections show that the duty of strict right requires us to regard others (i) as if they were *automata spiritualia* and (ii) as if no wrongs were imputable to them. Kant himself appears to endorse (i) in a passage in the *Vigilantius Ethics* on our duties to avoid "enmity, hatred, wrath, vengeance, etc.", which he groups together into the "duty to hate nobody":

to resist the offender...is a right of the person offended[.]...But anyone who seeks to avert unjustly perpetrated evil from himself treats and considers the other merely as a machine (*nur als eine Maschine*) that may be injurious to himself, and to which he assigns limits under which it cannot do him any harm. (VE 27:686)

Kant himself does not endorse (ii) in this passage, but he continues by suggesting that regarding the offender as a machine helps us avoid hatred:

So little...is a hatred of the offender's person connected with this, that it is...quite possible to make the resistance without hatred, although this is difficult for human nature; thus even in such resistance there is no enmity to the offender. (VE 27:686)

As Kant notes in the first quote from this passage, vengeance is a salient aspect of the hateful feelings often felt toward offenders, and vengeful attitudes often play a role in retributive attitudes. Kant holds that “no punishment...may be inflicted out of hatred” (MM 6:461), but in the *Anthropology*, he recognizes that vengeance and retribution are intertwined in human nature:

[H]atred arising from an injustice we have suffered, that is, the *desire for vengeance*, is a passion that follows irresistibly from the nature of the human being, and, malicious as it may be, maxims of reason are nevertheless interwoven with the inclination by virtue of the permissible *desire for justice*, whose analogue it is. (AP 7:270)

In the Vigilantius passages just quoted, Kant gives us a strategy for avoiding vengeance: strict right gives us a perspective on others as machines which allows resistance without vengeance. Kant does not employ this strategy to avoid retributive attitudes, but he should. Since nothing is imputable to machines (whether they be *materiale* or *spirituale*), considering people as machines is considering them as inappropriate subjects for imputation and retribution.

This account of the practical zero perspective raises a puzzle about *mens rea*, the “guilty mind” requirement in much of the criminal law which requires prosecutors to demonstrate that the accused acted with consciousness of misconduct to convict them of crime. Sarah Holtman argues that it is the notion of retributive proportionality expressed in Kant’s principle of equality

which provides a place for *mens rea* in Kant's philosophy of law ("Kant, Retributivism", 112). The practical zero perspective not only prescind from the principle of equality, but also requires us to regard the coerced as machines without obligations who cannot deserve to suffer. How can *mens rea* play a role in practices for responding to wrongdoers from this perspective? It depends on how we understand *mens rea* and why we think it is important. If we understand it as entailing desert of suffering, it cannot play a role. But if we understand it more austere, as a principle that requires us to index responses to agents who hinder rightful freedom to the motivational states that prompt their hindering behavior, then it can play a role. As discussed earlier, even the theoretical zero perspective allows us to reference mechanistically caused *Bösartigkeit* for purposes of explanation and prediction. An explanatory and predictive orientation to motivational states would suffice for giving agents' states of mind an important role in criminal law, for example, by distinguishing between accidental and intentional hindrances of rightful freedom. We often have grounds for predicting that agents who hinder rightful freedom intentionally will do so again, and we typically have no such grounds when such hindrances are accidental. This means that if the purpose of the criminal law is to coerce compliance with right, it has special reasons to focus on coercing the future behavior of intentional hinderers, and lacks such reasons with respect to accidental hinderers.

To understand Kant's texts, it is of course important to distinguish his own strict right perspective from the practical zero perspective this paper claims it entails. Some of the relationships between strict right and the other perspectives he acknowledges are clear. Strict right and theoretical zero both view subjects of coercion as machines without obligations, but theoretical zero does not include the obligations of coercers found in strict right. Strict right differs from the other practical perspectives in that it is a perspective of right rather than ethics,

and thus contains no ethical obligations. Practical infinity's bearing on both inner and outer action differentiates it from strict right, which bears only on outer action. Practical degrees and strict right are both perspectives we take on others rather than ourselves, and both regard others' freedom as constrained by the force of pathological incentives; however, practical degrees regards wrongdoers as violating obligations, while strict right does not.

Since strict right and practical equality are both perspectives on public law, their domains overlap, and their relationship is puzzling. One feature of strict right's strictness is its *complete externality* (MM 6:232). But another feature of strictness is the *exactitude* discussed earlier, in the practical equality section (MM 6:233, 6:332). Complete externality blocks appeals to others' obligations, and requires us to determine their choices with external sensible incentives, in a way that requires us to treat and consider them as machines. But as discussed earlier, Kant thinks *exactitude* demands the principle of equality, which indexes desert of punishment to *inner* wickedness, and this is ruled out by complete externality.

This contradiction may seem less problematic when considered in additional context, which provides the domain distinction between strict right and practical equality that previously seemed to be missing. Kant appears to think we ought adopt the strict right perspective on *potential* criminals, but we ought to switch to practical equality for *actual* criminals. We all begin with an equal right to coerce one another to comply with public law, but criminals *forfeit* this right, along with what Kant calls their "civil personality" (MM 6:283, 331). *Theory and Practice* provides helpful details:

whoever is subject to laws...can coerce any other...through public law. . . through which every other also resists him in like measure; but no one can lose this authorization to coerce . . . except by his own crime...[H]e cannot give it away of his own accord, that is, by a contract, and so bring it about by a rightful action that he has no rights but only duties[.] (TP 8:291–92)

Kant tells us that “beings that have only duties but no rights” are “human beings without personality (serfs, slaves)” (MM 6:241), and explains that it is “by the principle of retribution” that the convicted thief “must let [the state] have his powers for any kind of work it pleases (in convict or prison labor) and is reduced to the status of a slave” (MM 6:333). Kant’s view is thus that retribution justifies expelling actual criminals from the domain of reciprocal coercion which constitutes civil society.

Kant also appeals to the principle of retribution to justify generating the suffering which is the distal cause of the incentives of aversion meant to deter potential criminals from becoming actual criminals. As discussed above, Kant thinks punishing criminals for reasons of utility uses them as mere means (MM 6:332-3). The prohibition on using persons as mere means is fundamental to Kant’s ethics, and he seems to hold that only retributive punishment conforms to it. His implicit view is that the duty to coerce requires us to *use* retributively-justified suffering to generate proximal causes of fear, for example, by publicizing it effectively. This may explain Kant’s apparent endorsement of public hangings (AP 7:238).

Thus, Kant’s switch between the perspective of strict right for potential criminals and practical equality for actual criminals involves two moments: the crime retributively justifies jettisoning the criminal from the civil world where spheres of external freedom are equalized, as well as hurting him to generate the fear that mechanically coerces those who remain in the civil world to comply with law. The criminal is a means, but not a *mere* means, because he freely forfeited his place in the civil world and deserves to suffer.

These connections between coercion and retribution explain why Kant “mingles” right with ethics, despite claiming they are distinct—if right requires sensible incentives to motivate compliance, and only retributive justifications drawn from ethics can justify those incentives,

then right must depend on ethics. This reveals a straightforward way to make Kant’s view consistent: we can simply revise his claims about the distinctness of these domains of practical rationality. Below, I argue that this is a problematic resolution, because it is false that only retribution can justify the incentives required by right.

2. Reasons to adopt practical zero in public law

Part one argued that with a simple revision, Kant’s perspectives can be fit together consistently despite the varying quantities of retribution they prescribe. However, there is a widely-shared intuition that justifications for seriously harmful retribution bear a very heavy burden of proof, so consistency alone is not enough. Part two poses questions about Kant’s grounds for applying the perspectives, in order to argue that Kant’s account cannot bear this burden. This intuition is relevant for all retributive practices which can inflict serious harm, but the argument here focuses on practices in public law for responding to wrongdoers.⁸

In criminal law, the burden of proof intuition is expressed in broad support for the “beyond reasonable doubt” standard in the criminal court. Kant endorses an exceptionally strong and expansive version in the *Vigilantius Ethics*. He describes the standard the court must meet as “the greatest possible^t [*größtmöglichste*] moral and logical certainty [*Gewißheit*]” (VE 27:566), and claims it extends to the questions of (1) whether the crime was committed by the accused (whether “the man did it” [VE 27:567]), (2) whether he had a culpable “motive [for] the action” (VE 27:559), and also (3) whether he acted “with freedom”, emphasizing that “it is only when considered as a free being that he can be accountable” (VE 27:559)—legal imputation

⁸ Vilhauer (“Kantian Remorse”) offers a non-retributive account of conscience.

requires certainty that the crime was “chosen with free will” (VE 27:561). Objectors may doubt that Kant really means to require certainty about the metaphysics of free will to satisfy the burden of proof. However, the *Vigilantius Ethics* discussion of right incorporates a distinction between *homo noumenon* and *homo phenomenon* which appears identical to the distinction between the intelligible and empirical characters, so the text encourages us to take him at his word (VE 27:593). Since Kant’s theory of freedom depends on his practical philosophy, he appears to require certainty about practical philosophy as a whole to justify retribution in criminal law. This is requiring a lot, but it is not a bizarre requirement. In contemporary society, the most serious intentionally inflicted harms which are broadly (if often uncritically) regarded as susceptible to justification are the harms we inflict under the cover of criminal law and the law of war. The harms of criminal law are distinct from the harms of war in that society often contemplates inflicting them upon people who no longer pose threats because they have been captured and incarcerated. They are thus pragmatically optional (as it were) in a way that the harms of war are not. If we seek global and foundational certainty about justification anywhere in life, it makes sense to seek it here.

Now, as noted earlier, the practical epistemology of the second *Critique* and thereafter holds that it is not especially difficult to acquire certainty about the metaphysics of free will—the “fact of reason” lets us know the “cans” of freedom by inferring them from the “oughts” of practical reason. Kant thinks we can thereby attain certainty that people have the freedom necessary to deserve whatever retribution practical laws tell us to inflict. Can this practical epistemology meet the burden of proof necessary to justify practical equality in criminal law? Circularity threatens if we approach this question in isolation, since the practical epistemology allows anybody who claims certainty about the justice of practical equality to claim certainty

about the freedom necessary to justify it. Incorporating the distinction among the four practical perspectives provides more traction: can we really be certain that we ought to adopt the practical equality perspective in public law, rather than any of the other practical perspectives? Let us work with the notion of *very strong reasons for belief* in place of *certainty*, since it is hard to be confident about what certainty means in matters outside logic and mathematics if we step back a bit from Kant's texts. The question then becomes: do we have very strong reasons to adopt practical equality in criminal law, rather than one of the other perspectives? (This substitution should be unproblematic, since lacking extremely strong reasons presumably entails lacking certainty.) The need for very strong reasons for seriously harmful retribution implies that if a lesser quantity of retribution is a reasonable alternative to a greater quantity, and the quantum of difference is enough to be seriously harmful, we are obligated to go with the lesser. Let us compare practical equality to the other practical perspectives in light of this principle, starting with the most stringent.

Are there reasons to punish criminals according to practical infinity rather than practical equality? As discussed, Kant holds that punishment must target criminals' "*inner wickedness*" (MM 6:333). If their wickedness is infinite, why not punish it infinitely? Kant holds that the court does not punish in light of the infinite evil of evil maxims, but in light of the particular evil involved in the crime. Perhaps Kant infers from the particular external wrong to a particular inner action which grounds it, and represents that inner action as finite in a way that corresponds to the finite external wrong. Perhaps his policy is to regard all external wrongs as grounded in volitions to make exceptions to underlying moral maxims despite our knowledge of criminals' infinite wickedness. But *why* should this be his policy if we *know* that criminals are infinitely wicked and deserve infinite punishment? Why not do our best to inflict infinite punishment?

The human capacity for pain is finite in this life, but we could pursue infinite punishment as an ideal. We would not have to adopt a cruel attitude or disfiguring, humanity-dishonoring techniques—we could consider it our solemn duty to subject every shoplifter to maximally painful electric shock over the longest possible period. This would produce terrifically effective incentives of aversion. Kant’s reason *not* to do this appears to be based on a burden of proof argument:

We cannot pass any complete moral judgement on another, as to whether he is punishable or not before the divine judgement-seat, since we do not know his disposition. The moral dispositions of others are therefore a matter for God (CE 27:451).

We do not really *know* that criminals are infinitely evil, because we do not really know their dispositions. Our reasons to believe that criminals are infinitely evil are not strong enough to meet the burden of proof necessary to punish infinitely in public law.

Next, why should we adopt practical equality in public law rather than practical degrees? As discussed above, one of Kant’s reasons is that punishing different people in different ways for the same external actions, based on different degrees of imputation, would violate universality in the rule of law. But this is false, because we can have universal laws which incorporate degrees of imputation but apply equally to everyone, and arguably there are such laws today. Another reason is based on exactitude: Kant claims that only the principle of equality “can specify definitely the quality and the quantity of punishment”, and “all other principles are fluctuating...because extraneous considerations are mixed into them” (MM 6:332, also see VE 27:555, where Kant says punishing *gradatim* [by degrees] would “degenerate into mere arbitrariness”). But how can Kant’s own doctrine of degrees of imputation be an extraneous consideration? Kant insists that “[i]n every punishment as such there must first be justice” (2C 5:37), and the justice of punishment lies in its being “an immediately necessary consequence of

the morally bad act” (VE:27:552). Kant explains this idea in the *Metaphysics of Morals* with the claim that “[t]he rightful effect of what is culpable is punishment” (MM 6:227). The passage in which he explains that culpability comes in degrees appears just three short paragraphs later. Incorporating degrees of culpability into criminal law need not yield fluctuating arbitrariness—on the contrary, calibrating punishments to criminals’ degrees of culpability would yield a *more* exact correspondence to desert, since, as Kant acknowledges, “*duo cum faciunt idem, non est idem* [when two do the same, it is not the same]” (VE 27: 563). Reducing retribution when culpability is reduced is a matter of responding to reasons of justice, not consequentialist reasons—it is not a matter of “reduc[ing] its amount by the advantage it promises”, an approach which we saw Kant critique in the discussion of practical equality above. Further, Kant himself acknowledges constraints on the exactitude possible under the principle of equality which could also reasonably prompt charges of arbitrariness. In some cases, crimes “cannot be punished by a return for them because this would be . . . itself a punishable crime...for example, rape” (MM 6:363). In other cases “differences in social rank” may “not allow the principle of retribution, of like for like”—he mentions a fine which may be negligible for the wealthy, implying that it could nonetheless be damaging for the poor (MM 6:332). Enslaving thieves (MM 6:333) and burning wives for trying to poison their husbands (CE 27:452) are also less-than-transparent examples of exact equality, though Kant does not acknowledge a concern in these cases. Why should we think exactitude is a problem for practical degrees if it is not a problem for practical equality?

It is also significant that practical degrees involves a concept of freedom that is different from the concept of freedom Kant appeals to in practical infinity and equality—one where the *difficulty* of doing what one ought to do constrains culpability. If we take the practical

epistemology of freedom seriously, then the practical degrees perspective implies that the freedom we know through practical reason is sometimes a constrained freedom. What principles of practical epistemology could allow us to know freedom as constrained in some cases, but as unconstrained when we contemplate grievous harms like execution and enslavement, where the burden of justification is at its highest? Kant offers no answers. Since practical equality prescribes violence and practical degrees does not, serious harm is at stake in the difference. Kant does not consider a burden of proof argument here, but he should, so we ought to prefer practical degrees to practical equality in criminal law.

Next, why should punishment in public law be governed by practical equality rather than the practical zero perspective which is entailed by strict right according to the arguments above? Both apply to public law, and while the practical epistemology infers criminals' unconstrained freedom from practical equality, it infers *nothing* about criminals' freedom from practical zero, because it contains no obligations on the part of criminals. Further, if the practical epistemology says that the *only* way to infer freedom is from obligations, then claims about criminals' freedom are *disallowed* from the practical zero perspective. Why should practical equality epistemologically trump practical zero in public punishment? Since practical equality prescribes like-for-like retribution and practical zero prescribes none, it seems clear that the burden of proof favors practical zero. Before we draw this conclusion, however, we must revisit Kant's implicit assumption that right must rely on the ethics of retribution to justify the incentives needed to coerce compliance with law. If this assumption is true, then perhaps it is a confusion to think of practical zero as an *alternative* to practical equality. But this assumption is false. Strict right itself can generate sufficient incentives of aversion, without drawing on retribution. Two such incentives which would together be sufficient are (i) incentives of fatigue, and (ii) incentives of

deterrence from preventative incarceration justified via non-retributive contractualism, rather than via retributivism or utilitarianism.

(i) The fundamental idea animating Kant's justification of coercion is that hindering hindrances to rightful external freedom is necessarily rightful. One way to describe the intuition which makes this compelling is that when somebody *pushes* us in a way that obstructs us in doing what we have a right to do, we must be entitled to *push back* enough to continue with our rightful activity. Kant himself suggests this picture when he says that the "law of a reciprocal coercion" can be constructed "by analogy with . . . bodies moving freely under the law of the equality of action and reaction" (MM 6:232). A pushing match can generate incentives of aversion sufficient to prevent crime all on its own: if rightful pushers push back persistently enough, wrongful pushers eventually tire and cease their attempts. This produces rightfully justified incentives of fatigue rather than retributively justified incentives of fear. While Kant seems only to contemplate the latter, the former can also be effective, and their justification does not implicate ethics in the way retributive justification does. The former also offer a precise notion of exactitude: we push back against criminals with exactly the force necessary to stop crime.

Kant may think there is a kind of pragmatic necessity in punishing retributively after crimes have been committed, given present constraints on our capacity to push back enough to stop all crime. But social and technological innovations could allow us to push back better. Perhaps we could put careful, compassionate rightful pushers on patrol in sufficient number to fatigue every would-be criminal. Perhaps we could design artificially-intelligent and constitutionally-adroit rightful pushers: soft, ubiquitous nanodrones that know when to apply gentle but fatiguing pressure, or brainwave hats that know when to induce somnolence. At

present, it is hard to imagine trying to stop all crime without oppressive surveillance and policing which stops not only crime but also some rightful activity which flawed authorities misclassify as crime. Giving flawed authorities better technologies for coercion would lead to great injustice. However, to the best of our knowledge, this is a contingent and temporary constraint. Strict right gives us a duty to imagine more inventively, and design authorities as well as technologies that can coerce without undermining right.

(ii) Society would be reluctant to rely on incentives of fatigue alone while developing innovations that made them sufficient for right. While we were innovating, we would fail to prevent many crimes. From the perspective of strict right, this is the coercers' failure to do their duty, not the criminals' failure, but leaving criminals unsanctioned would threaten the domain of right with a collapse into the state of nature. Even if we reject retributivism, we would seek a justification for preventatively imprisoning at least some criminals, especially violent criminals likely to reoffend, within prison walls that can do the "pushing back" on our behalf. We would also hope to justify prison conditions that are not too comfortable—if prison conditions provide no incentives of aversion, they could not deter, and could become additional incentives to commit crime. Can we have such a non-retributive justification without appealing to a utilitarianism that Kantianism cannot allow? We can, if we appeal to a contractualism which embodies another feature of the revised strict right proposed here—that criminals cannot be jettisoned from the civil domain of reciprocal coercion. As explained above, Kant appeals to retributivism to excise criminals from this domain. He thinks "criminal[s] cannot possibly have a voice in legislation", since "the legislator is holy", and the fact that we are "capable of crime" is irrelevant in consenting to the social contract which articulates the framework of reciprocal

coercion (MM 6:335). If the appeal to retribution is blocked, criminals must have a voice, and must be able to rationally consent to the social contract.

We can build a practical zero perspective which includes criminals in rational and fair contract bargaining by selecting principles of criminal law through a modified form of Rawlsian original position deliberation (OPD).⁹ Rawls himself applies OPD to distributive justice, not criminal justice, but it can be extended to criminal justice if modified. Standard OPD uses maximin reasoning to maximize welfare for the worst-off members of society. Maximin reasoning cannot be directly applied in criminal justice, because of the way criminals and potential victims compete: assuming that sanctions on criminals deter, diminishing sanctions increases danger for potential victims, and increasing sanctions diminishes danger for potential victims. Given the inapplicability of maximin reasoning, a plausible alternative for making bargaining *fair* is to assume we are *equally likely* to be sanctioned criminals and potential victims when the veil is raised. What criminal sanctions would we choose under this assumption?

Deliberators' fear of sanctions would make them prefer a society which does not need to sanction criminals because it stops all crime, so they would invest in developing the innovations in (i) above. They would also invest in incentives that make the lives of the law-abiding more attractive: better jobs, education, and public services, and free voluntary therapy and social support for people at risk of committing crimes. This would provide incentives of "allurement" which "invite" people to follow the law (MM 6:218-9). Kant thinks there is no role for

⁹ Some remarks in this paper are drawn from Vilhauer ("Persons, Punishment" and "Kant's Mature Theory"), which explain this view of punishment in more detail. It has much in common with Dolovich ("Legitimate Punishment"), though Dolovich's account has retributive elements inadmissible in practical zero.

incentives of allurements in public law, but deliberators in fear of sanctions would think differently.

Deliberators would also fear victimization, however, and would endorse preventative detention of (at least) violent criminals likely to reoffend, to avoid the decay of reciprocal coercion and the state of nature to which it leads. Since deliberators assumed they might find themselves imprisoned when the veil is raised, they would choose humane prison conditions with reforms such as continual parole review and better opportunities for education and meaningful work. But they would not make prison conditions too comfortable—they would want conditions to deter exactly enough to prevent them from becoming a new incentive to commit crime. Conditions would not have to be intrinsically unpleasant to deter, just aversive relative to life outside prison. Since deliberators would fund alluring jobs, education, and social services outside prison, life would get better outside prison, so it could also get better inside prison while still deterring. Objectors may think it would be irrational to consent to a social contract if we assumed a 50% chance of imprisonment, since the state of nature would be preferable to imprisonment. But the reformed prison conditions demanded by deliberators would be preferable to the state of nature and its war of all against all.

Since deliberators' motivation for allowing imprisonment would be to avoid the state of nature, it would be consequentialist. But this consequentialist motivation only has normative significance because it derives from rational consent, so this justification of imprisonment is fundamentally deontological despite eschewing retributivism. The imprisoned are used as *means* in this framework, but the grounding in rational consent implies that they are not used as *mere* means. Further, imprisoned people reciprocally use people outside prison as means, coercing them through the social contract to provide the resources necessary to improve life inside prison.

Thus far, part two of this paper has argued that the burden of proof in criminal law implies that practical equality is preferred over practical infinity, and practical zero and practical degrees are preferred over practical equality. It has not yet argued that practical zero is preferred over practical degrees. Appealing to the special burden of proof for serious harm has less force here, since Kant advocates appealing to practical degrees only when punishments less severe than those of criminal law are at issue. But it still has considerable force. As mentioned above, Kant thinks ostracism can sometimes be prescribed as punishment from the degrees perspective, and ostracism could conceivably be adopted as a criminal sanction. Ostracism has much in common with imprisonment, and while deliberators in our modified OPD would endorse imprisonment, they would insist on continual parole review, so prisoners could rejoin society when they no longer pose threats. Permanent ostracism is seriously harmful, and since practical zero imprisonment is a reasonable alternative, we ought to prefer it. However, practical degrees and practical zero do not always prescribe different treatment, since there could be no distinction in practice between imputing in the vanishingly small degree which marks the lower boundary of degrees of imputation and not imputing at all, as practical zero requires. But while the burden of proof gives us reasons to prefer practical zero to practical degrees only some of the time, it never gives us reasons to prefer practical degrees to practical zero. Thus we can be sure not to impose unjustified retribution if we adopt practical zero rather than practical degrees in criminal law.

Conclusion

To conclude, let us consider a general but important objection to this paper's approach. It may seem that, despite the tensions in Kant's own account of retribution and the reasons provided for the revision advocated here, this revision foists a distorting non-retributivist approach to criminal justice onto Kantianism which is alien to its spirit and cannot provide

insight into its nature. I think this objection is mistaken. Consider a passage from the first *Critique*'s discussion of referring imputation to the empirical character, part of which was quoted earlier. Kant argues that, since we are "not acquainted" with the intelligible character,

the real morality of actions (their merit and guilt), even that of our own conduct, therefore remains entirely hidden from us. Our imputations can be referred only to the empirical character. How much of it is to be ascribed to mere nature and innocent defects of temperament...no one can discover, and hence no one can judge it with complete justice. (A551n/B579n)

It is natural to read Kant as applying a burden of proof argument here: the unknowability of the intelligible character means that we cannot meet the burden of proof necessary to judge with complete justice, thus we ought to refer imputation to the mechanistic empirical character. As discussed above, this can only mean dissolving imputation into merely causal responsibility and undermining the justification of retribution. Kant's focus in discussing the empirical character is to explain the theoretical perspective on human agency. But since the world contains no "oughts" from the theoretical perspective, following the moral principle implied by the burden of proof requires a practical perspective. Thus Kant is implicitly adopting a practical perspective in this remark. He makes this remark before he develops the practical epistemology of the second *Critique*, and after he develops this epistemology, he never again suggests a practical perspective which depicts people mechanistically *and* undermines the justification of retribution. But as argued here, even by Kant's own lights, strict right *does* imply a practical perspective which depicts people mechanistically, and while Kant does not suggest that it undermines retributivism in criminal law, he ought to acknowledge that it does. If we correct this error, as this paper has sought to do, then the practical perspective strict right entails has much in common with the one Kant implicitly adopts in this first *Critique* remark. Strict right becomes a localized infolding into the practical philosophy of the theoretical perspective's

implications for retribution. It short-circuits justification of some of the most serious harms humans seek to inflict on each other, and it does so without globally undermining our perspective on human beings as free, morally responsible, and deserving. It can thus be understood as a development of a perspective Kant himself suggests early in the critical philosophy—a Kantian road not taken rather than an alien imposition.

Acknowledgements

Thanks to Matthew Altman, Paul Guyer, Susan Meld Shell, Marcus Willaschek, Allen Wood, and the reviewers and editors.

Bibliography

- Allison, Henry E. *Kant's Conception of Freedom: A Developmental and Critical Analysis*. Cambridge: Cambridge University Press, 2019.
- Blöser, Claudia. "Degrees of Responsibility in Kant's Practical Philosophy". *Kantian Review* 20, no. 2 (2015): 183–209.
- Dolovich, Sharon. "Legitimate Punishment in Liberal Democracy." *Buffalo Law Review* 7, no. 2 (2004): 307-442.
- Fabbianelli, Faustino. "Kant's Concept of Moral *Imputatio*". In *Reading Kant's Lectures*, edited by Robert R. Clewis, 200-220. Berlin: Walter de Gruyter, 2015.
- Holtman, Sarah. "Kant, Retributivism, and Civic Respect". In *Retributivism: Essays on Theory and Policy*, edited by Mark D. White, 107-128. Oxford: Oxford University Press, 2011.
- Palmquist, Stephen R. *Comprehensive Commentary on Kant's Religion within the Bounds of Bare Reason*. Oxford: Wiley-Blackwell, 2016.
- Vilhauer, Benjamin. "Persons, Punishment, and Free Will Skepticism". *Philosophical Studies*, 162, no. 2 (2013): 143-163.
- Vilhauer, Benjamin. "Kant's Mature Theory of Punishment and a First *Critique* Ideal Abolitionist Alternative." In *Palgrave Kant Handbook*, edited by Matthew C. Altman, 617-642. New York: Palgrave Macmillan, 2017.
- Vilhauer, Benjamin. "Kantian Remorse with and without Self-Retribution". *Kantian Review* 27, no. 3 (2022): 421-441.

Willaschek, Marcus. "Which Imperatives for Right? On the Non-Prescriptive Character of Juridical Laws in Kant's Metaphysics of Morals". In *Kant's Metaphysics of Morals: Interpretative Essays*, edited by Mark Timmons, 65-87. Oxford: Clarendon Press, 2002.