

Persons, Punishment, and Free Will Skepticism

Philosophical Studies, January 2013, Vol. 162, No. 2, pp. 143-163.

Benjamin Vilhauer, William Paterson University of New Jersey

1. Introduction

Free will skeptics are typically eager to point out that their view requires them to reject retributivism. (Free will is the kind of control which is necessary for moral responsibility. Free will skepticism is the view that human beings do not have free will. Retributivism is the view that punishment is wholly justified by the fact that it inflicts on criminals the suffering they deserve for their actions.¹) Free will skepticism is incompatible with retributivism because action-based desert depends on moral responsibility. In other words, we only deserve to be treated in particular ways based on our actions if we are morally responsible for them.²

It is not hard to understand why free will skeptics might be eager to point out this incompatibility. When looked at in one way, retributivism can seem to be hardhearted and even cruel. Retributivists typically acknowledge that punishing criminals probably has good effects on society, for example, by helping to control crime. But they hold that these effects are merely fortunate side-effects of punishment, not part of what makes it just. In their eyes, even if punishment only served to make criminals suffer and had no good effects elsewhere in society, it would be no less justified. Many people find this dimension of retributivism disturbing even if they think that we have free will. This makes it is natural for free will skeptics to think that the anti-retributivist implications of their view are not a grave liability. These implications may even be a selling point in some cases.

But there is another dimension of retributivism which should make free will skeptics uneasy about rejecting it. Retributivists commonly hold that theirs is the only justification of

punishment which avoids an objection about using persons, which is typically leveled against utilitarians.

Utilitarians hold that punishment is justified not because criminals deserve it, but instead because it reduces overall suffering in society. They hold that criminals' suffering is intrinsically bad, and can only be justified if it prevents a greater amount of suffering than what is inflicted on the criminal. This approach can make utilitarianism seem benign, at least by comparison with retributivism. And since utilitarians do not claim that criminals deserve to suffer, free will skeptics can accept utilitarianism without any trouble. This may initially sound like a simple and happy resolution for free will skeptics. But if our sole justification for punishment is that it reduces overall suffering in society, then there is a clear sense in which we are using the people punished as mere tools that we manipulate in order to benefit others. Further, suppose it turns out that we get the most deterrence for the least punitive pain with practices that are startlingly instrumentalizing: imposing maximally painful punishments, weakening or violating due process, and framing non-criminals when it is more effective than punishing real criminals. If these practices turn out to give us the lowest ratio of suffering-caused-to-suffering-prevented, then utilitarianism must endorse them.³ Retributivism's refusal to count the harm-reducing consequences of punishment as part of the justification of punishment immunizes it against this "mere means objection".⁴

When retributivism is seen in this light, it can appear to be a bastion of the Kantian principle that people must never be treated as mere means. Some free will skeptics may doubt that this notion is worth defending, because they may think that the sort of freedom possessed by people who are not treated as mere means is not morally significant unless it is backed up by free will. But this is a mistake, even if Kant himself at times seems to think that it is true.⁵ Free will

skeptics would do themselves a service if they found a way to justify punishment which does not imply that it is all right to treat people as mere means. I think that retributivists are correct to hold that their view has an important advantage over utilitarianism when it comes to replying to the mere means objection. But if these two justifications are not the only options, then we cannot conclude without further argument that there is not some other justification which fares as well as retributivism. If that other justification is compatible with free will skepticism, then free will skeptics can reply to the mere means objection.

Kant himself is of course a notorious retributivist, and his views on punishment are complicated and arguably contradictory at points. I will not attempt an exegesis of them here.⁶ Instead, I want to look at punishment in terms of a central Kantian guideline on how we avoid treating people as mere means: we must treat them as they would rationally consent to be treated. If this is the right approach, then to mount a defense against the mere means objection, a justification of punishment would have to be able to explain why somebody would rationally consent to punishment. This means that if retributivism really were the only justification of punishment which could mount a defense against the mere means objection, then it would have to be the only one which could explain why someone would rationally consent to punishment.

This might be true if there were some special connection between the concept of rational consent to punishment and the concept of action-based desert upon which retributivism is founded. Is it plausible that such a connection exists? It depends upon how we understand the idea of rational consent. One popular approach to rational consent is the "actual consent" account. The central idea in all varieties of this account is that people consent to being treated in a particular way when they act in a particular way. Typically, this is not explicit, by way of a speech act, but instead implicit, by way of an act that implies consent. When applied to

punishment, the idea is that when the wrongdoer harms someone, he must acknowledge (on pain of irrationality) that he has implicitly consented to be treated in this way himself, so that someone can justifiably do punitive harm to him in return. I am not sure how plausible this story is, but for the moment let us suppose that it makes sense. It might be argued that the wrongful action could only constitute genuinely binding consent of the sort that would legitimize punishment if the wrongdoer is morally responsible for the action. But this is the same relation between agents and actions that is supposed to be sufficient for action-based desert. So it might be argued that an action can only constitute consent to punishment if the agent can also deserve punishment based on that action. If this is right, it would support the idea that there is a special connection between action-based desert and actual consent to punishment, and it might well be the case that any justification of punishment that has the conceptual resources needed to explain actual consent to punishment has the same basic structure as retributivism.

But the actual consent account is not the only account of rational consent, and it is by no means clear that it is a satisfactory account. (For example, as has often been observed, the claim that people who wrongfully harm others implicitly consent to be harmed in return is a bit mysterious, to say the least.⁷⁾ Another popular account, which also happens to be more congenial to free will skepticism, is the “hypothetical consent” account. It does not seek to identify actions in the actual world which constitute consent. Instead, it asks whether an agent would have consented to be treated in some particular way if she had thought rationally about it. If the answer is “yes”, then the rational consent requirement is considered to be satisfied. Obviously a great deal turns on how we explain what it is to think rationally about things, and I will discuss that later. For the moment, I want to focus on the following point. When applied to crime and punishment, this view says that it is legitimate to punish a criminal so long as a given

counterfactual conditional obtains: if the criminal had thought rationally about things, then he would have consented to be punished for his crime. It does not claim that what the criminal has actually done constitutes consent, only that it constitutes a crime. To make the same point in different terms, we might say the following. A hypothetical consent justification needs an anchor in the actual world (the crime that the criminal has committed), but the claim it makes about consent hooks into a world that may be merely possible, a world in which the criminal thinks rationally about things.⁸ The justificatory contribution made by the idea of consent does not depend upon what the criminal has done in the actual world.

For this reason, the concepts of rational consent and action-based desert cannot overlap on the hypothetical consent account in the way they may on the actual consent account. On the actual consent account, both the crime and the consent are constituted by the same (actual) action, so retributivists have room to argue that claims about consent depend upon moral responsibility in the same way that claims about action-based desert do. But on the hypothetical consent account, claims about consent come in the form of counterfactual conditionals. Since we can only be morally responsible for what we actually do, the justificatory force of claims about hypothetical consent cannot depend upon an assertion of moral responsibility. So retributivists cannot argue that we need the same conceptual resources to explain hypothetical consent that we need to explain action-based desert.

In other words, while it seemed plausible on the actual consent account to suppose that consent can only be legitimately binding if one is morally responsible for the action which constitutes it, and to conclude that we must appeal to the same metaphysics of moral responsibility to support the notion of consent that retributivists require to explain action-based desert, no parallel claim can be made about the hypothetical consent account. On the

hypothetical consent account, there is nothing the criminal actually does which constitutes consent, so there can be no requirement to be morally responsible for it, because one can only be morally responsible for what one actually does. This means that free will skeptics need not worry that a metaphysics of moral responsibility is required to support the role played by consent in the hypothetical consent account. This implies that if free will skeptics can build a hypothetical consent account which does not appeal to moral responsibility to explain the significance of its anchor in actuality (the crime itself) then they ought to be able to avoid entanglement with moral responsibility altogether, since no further problems about moral responsibility can be raised by the notion of consent (since it is a mere hook into the possible on this account). The upshot is that, from the perspective of hypothetical consent, the claim that retributivists are uniquely qualified to explain rational consent to punishment is much less plausible.

Retributivists might reply that the metaphysics of moral responsibility is still crucial for the hypothetical consent account, because the hypothetical account must assume that agents *could have* rationally consented if they did not actually do so, and this implies that a possible world in which they rationally consented was *accessible* to them. Retributivists might claim that the metaphysics of accessible alternatives depends on the concept of free will, which is one of the conditions of action-based desert. They might claim that this shows that the hypothetical consent approach still depends on the conceptual resources of retributivism.

The problem with this reply is that if we look at how the hypothetical consent approach is typically formulated, it is pretty clear that it does not imply accessibility. That is, the hypothetical consent approach rests on the claim that the criminal *would have* consented if he had thought rationally about it, but this does not imply that the criminal *could* have thought

rationally about it. Retributivists might claim that even if the typical formulation does not imply accessibility, the hypothetical consent approach is not a plausible reply to the mere means objection unless we assume accessibility. That is, retributivists might claim that the fact that someone would have consented to punishment if he had thought rationally about it would only support the claim that we can punish him without using him as a mere means if he had it in his power to think rationally about it. But this is not *prima facie* true in any obvious way, and there are clear cases where the question of hypothetical rational consent matters despite the fact that agents are unable to think rationally. Suppose a cult has kidnapped someone and then brainwashed him into accepting his own enslavement, and we force him to undergo deprogramming therapy to free him from the brainwashing. It is obviously relevant to point out that he would have consented to deprogramming if he had thought rationally about it, but it is just as obvious that as things were in the actual world, he could not have thought rationally about it. The question of whether a possible world in which he thought rationally was accessible to him is simply not to the point here. Given this, the burden of proof would seem to be on retributivists to show that things are different with respect to punishment. If the arguments I will present later in the paper are correct, this burden cannot be met.

Some will no doubt still feel that the line of inquiry I am proposing is too confused to be worth pursuing. After all, prominent retributivists, including Jeffrie Murphy, have proposed hypothetical consent responses to the mere means objection. These often rely on social contract theory. Murphy proposes that we justify punishment by asking what principles of punishment would win our consent in the Rawlsian original position. In other words, he thinks that we ought to understand what it would be for criminals to think rationally about punishment in terms of original position deliberation: the principles of punishment to which criminals would rationally

consent are the ones they would choose in the original position. Am I claiming that Murphy's approach to justifying punishment is not a fundamentally retributivist approach? After all, the purpose of original position deliberation is to model our intuitions about fairness and respect for persons. Do we not deserve to be treated fairly and with respect? If we can claim to deserve these kinds of treatment, and if justified punishment is supposed to respect this desert claim, then how can this be anything other than a fundamentally retributivist perspective on punishment?

I think it is quite correct to see a connection between how we would rationally consent to be treated and how we deserve to be treated. But I think that an interesting account of this connection can be given without appealing to the notion of action-based desert. Many retributivists mistakenly assume that desert is monolithic. That is, retributivists assume that action-based desert is the only kind of desert. But there are also personhood-based desert claims, that is, desert claims which depend not on facts about our actions, but instead on the more abstract and general fact that we are persons. (I do not mean to claim that there are only these two kinds of desert, but rather that there are at least these two kinds.) Since personhood-based desert claims do not depend on facts about our actions, they do not depend on moral responsibility, so free will skeptics can appeal to them just as well as retributivists.

Retributivists must acknowledge that personhood-based desert exists, because we can claim to deserve due process, and as I will argue in section two, this desert claim is personhood-based. In sections three and four, I will go on to argue that personhood-based desert can be the kernel of a complete justification of punishment which can stand on its own, without any retributivist or consequentialist premises. What people deserve based on the mere fact of their personhood is to be treated as they would rationally consent to be treated if all they had in view was the mere fact of their personhood. We can work out the implications of this view by taking

a lead from Murphy and choosing principles of punishment in the Rawlsian original position, so long as we are careful not to smuggle in the assumption that it is under our control whether we end up as a criminal or as a law-abiding citizen once we raise the veil of ignorance. If this works, then it implies (among other things) that justifying punishment through original position deliberation is not, after all, a fundamentally retributivist approach to punishment.

2. The Distinction between Action- and Personhood-Based Desert

The purpose of this section is to argue that there is an important distinction between action- and personhood-based desert claims, and to point out some personhood-based desert claims that retributivists should be loath to disregard.⁹ A “desert base” is whatever grounds a desert claim. Many commonplace desert claims are based on actions, and some hold that all desert claims are based on actions.¹⁰ Examples of action-based desert claims are claims about praise and blame, and the Lockean claim that we come to deserve property when we "mix our labor" with objects. Retributive desert claims are also examples. That is, retributivists hold that criminals deserve hard treatment based on their criminal actions. There is broad agreement that an action can be a desert base only if an agent is morally responsible for that action. If this is right, then free will skeptics must hold that action-based desert claims (and by extension retributive desert claims) are never legitimate. But I think that there are some desert claims which cannot plausibly be supposed to be action-based. Some are based on the mere fact of being a person.¹¹ For example, when we claim to deserve respect, not to be used as mere means, to be given access to our rights, and to be granted due process of law, these claims are not based on facts about our actions, but instead on the fact that we are persons. Free will skepticism does not undermine personhood-based desert in the way it undermines action-based desert, so free will skeptics can endorse personhood-based desert claims.

Saul Smilansky is a notable critic of the idea that there is a fundamental distinction between action- and personhood-based desert. He holds that all non-trivial desert claims are action-based ("responsibility-based", in his terminology). Smilansky accepts that there is a minimum level (a "baseline") of well-being which persons deserve just because they are persons. But he argues that when it comes to explaining why people who fall below that baseline deserve special treatment, our explanation must be based on their actions.

To illustrate his claims, Smilansky provides an example about the residents of a town whose air has been seriously polluted by a chemical factory. He accepts that there is no "responsibility-base", i.e. no action-base, for the "desert of fresh air". That is, the residents can legitimately claim to deserve fresh air just because they are persons. But when they fall below the fresh air baseline because of the pollution, the mere fact that they are persons cannot explain why they deserve (e.g.) compensation. To explain why the residents deserve special treatment, we must base our claims on facts about their actions: "since the people of the town are *not responsible* for the pollution (they have done *nothing to deserve it*), they *deserve* e.g. compensation" (1996, 159). According to Smilansky, the residents' personhood-based desert claims play no role in the explanation of why they deserve special treatment, and in this sense, the personhood-based claims are trivial.

Smilansky holds that the relationship between personhood- and action-based desert which he points out in the pollution example exists in all situations in which we are inclined to appeal to personhood-based desert claims. If he were correct, then personhood-based desert claims could not be expected to do very much philosophical work when decoupled from action-based desert claims. But the fact that one has done nothing to deserve to fall below the baseline can only be the correct explanation of why one deserves special treatment if it is conceivable that one

could do something that would make one deserve to fall below it. In the pollution case, we can imagine things the residents might do to deserve to fall below the baseline, for example, burning too much of a polluting fuel. But in the cases of greatest interest for the present paper, there is nothing anyone could conceivably do to fall below the relevant baseline. Consider another of Smilansky's examples of baseline desert: "everyone should be treated as innocent until proven guilty"(2000, 44). There is nothing anyone could conceivably do to deserve not to be treated as innocent until proven guilty. Someone who was not treated as innocent until proven guilty could legitimately claim to deserve special treatment. But she could not explain this claim by pointing out that she had not done anything to deserve not to be treated as innocent until proven guilty, because there is nothing she could conceivably have done to deserve not to be treated as innocent. So in this case it is not true that her claim to deserve special treatment is action-based. In other words, this demonstrates that personhood-based claims to deserve to be treated as innocent until proven guilty do not have to be backed up by action-based desert claims to be substantive. The same can be seen to hold for the desert claims we make with respect to other customary aspects of due process, for example, our desert of equal treatment before the law, trial by a jury of peers, competent legal representation, etc. That is, there is nothing one could conceivably do to make it the case that one did not deserve these things. This also holds for our more abstract personhood-based claim to deserve not be used as a mere means.

At this point in the argument, retributivists might concede that I have demonstrated that particular personhood-based desert claims can be substantive even when they are not backed up by particular action-based desert claims. But they might insist that personhood-based desert claims depend upon action-based desert in a more general way. That is, they might claim that personhood-based desert claims can only be made by agents who have the capacity to be morally

responsible for their actions (that is, the capacity to act in ways that ground action-based desert claims). Or, more succinctly, they might claim that personhood implies moral responsibility.

I cannot discuss this argument in detail here, though I do so elsewhere.¹² In my view, anybody who is not already committed to compatibilism ought to think that the claim that personhood implies moral responsibility is pretty dubious. What I mean is that people not already committed to compatibilism typically think that the claim that human beings possess the free will needed for moral responsibility is metaphysically loaded. They think that it is quite conceivable that the metaphysical facts about the world might rule out moral responsibility. The claim that persons exist, on the other hand, seems metaphysically innocuous by comparison. Claims about personal identity or the nature of consciousness may become metaphysically loaded quite easily, but the more basic claim that persons exist seems unproblematic unless one seriously entertains skepticism about other minds. From this perspective, it seems odd to think that something metaphysically innocuous like the existence of persons should imply metaphysically loaded claims about free will.

Diehard compatibilists will not be moved by the argument just made, but they might consider the following points about Frankfurtian "wantons" (Frankfurt 1971). From the perspective of this paper, Frankfurt sets too high a standard for personhood. That is, wantons are clearly persons in the sense relevant for the personhood-based desert claims at issue in this paper. Despite their lack of second-order desires, wantons deserve not to be used as mere means, and if accused of a crime they would deserve due process. But Frankfurt is correct that wantons are not morally responsible for their actions. Therefore personhood, in the sense at issue here, does not imply moral responsibility.

The view that persons must be morally responsible for some of their actions is admittedly not uncommon. But I think this view is largely based on an assumption that certain properties which pretty clearly are essential to personhood, such as the capacity to do things for reasons (which wantons have), imply the capacity for moral responsibility (which wantons lack). I have never seen an argument for this view which does not include controversial compatibilistic or libertarian premises, and it seems to be a straightforward matter for free will skeptics to draw a line between the kind of control involved in doing things for reasons and the kind of control necessary for moral responsibility.

3. Personhood-Based Desert and the Original Position

The concepts of personhood-based desert and rational consent seem to dovetail nicely. Respecting people's personhood-based desert claims implies treating them in a way they can rationally consent to be treated. That is, we can make personhood-based desert claims not to be treated as mere means, and not treating people as mere means implies treating them as they can rationally consent to be treated. So a justification of punishment which respects personhood-based desert claims must be able to pass the rational consent test. The implication probably holds in the other direction as well: treating people as they would rationally consent to be treated probably implies respecting personhood-based desert claims.

The idea of the social contract can be used to model rational consent. If it would be rational to choose to enter into a social contract with a given institutional structure, then it makes sense to see that institutional structure as one to which we could rationally consent. My goal here is to model rational consent based on personhood- but not action-based desert. It may be that not all approaches to social contract theory fit this bill. The Rawlsian social contract seems

specially suited to the purpose, because original position deliberation has the effect of “filtering out” action-based desert claims.¹³

The original position is an imaginary epistemic standpoint in which the basic principles governing society are chosen. It is formed by drawing what Rawls calls a "veil of ignorance" between the people who make up a society and all their particular characteristics. For example, in the original position, I cannot know whether I am among the most or least wealthy or intelligent, or what my religion, ethnicity, sex, or sexual orientation is. I also cannot know what patterns of action I exhibit, for example, whether I am among the most or least industrious. So I must choose principles to govern society without knowing what my role in society will be, and without knowing how my particular characteristics will influence that role. Rawls thinks this is the fair way to choose the basic principles of society because he thinks we do not deserve to have our particular characteristics, so we should not be advantaged or disadvantaged by them.¹⁴

In conditions of such radical ignorance, Rawls thinks it is both natural and just to worry most about what it would be like to be among the worst-off members of society. Rawls concludes that we would choose principles that ensured that everyone possessed equal political rights, that everyone's basic needs for food, education, health care, etc., were satisfied, and that inequalities in wealth were allowed only insofar as they served to make the poorest richer than they would have been in the absence of those inequalities. Rawls calls the last bit, about the conditions under which economic inequalities are acceptable, the "difference principle".

It is important that Rawls includes facts about how we act among the particular characteristics which are to be hidden from us by the veil of ignorance, and that he is motivated to do this by a kind of skepticism about desert. It would be overstating the case to say that Rawls is anything like a thoroughgoing skeptic about free will or action-based desert. But it seems

reasonable to claim that Rawls thinks that action-based desert claims should not play a role in justifying the basic principles of society, and that he has designed the original position partly with an eye toward preventing them from doing so.

Rawls does not talk about personhood-based desert. But I think we can interpret his view in terms of personhood-based desert without imposing anything alien on it. There is a straightforward sense in which the mere, general fact of one's personhood is what remains when all the particular facts about oneself are obscured by the veil of ignorance. Further, one function of the veil of ignorance is to prevent the particular facts from serving as desert-bases, but to allow personhood to continue to serve as a justificatory foundation. From the perspective of this paper, it makes sense to characterize the way in which it serves as a justificatory foundation by saying that it supports personhood-based desert claims. Rawls' own preferred terminology is to describe claims we can make based on personhood alone as entitlement claims rather than desert claims. But this seems to be merely a terminological distinction.¹⁵

If this line of thought makes sense, then the Rawlsian social contract can be used to model what it would be rational to consent to if we had in view personhood- but not action-based desert claims. This means that we can use original position deliberation to develop a justification of punishment which has a ready reply to the mere means objection, but which can nonetheless be endorsed by free will skeptics.

I should note that Rawls himself rejects the idea of choosing principles of punishment in the original position. In section 48 of *A Theory of Justice*, Rawls implies that original position deliberation can only be applied to distributive justice, not to punishment, because the purpose of punishment is to sanction the “bad character” of criminals, while questions of character are morally irrelevant to distributive justice. Many commentators have thought this asymmetrical

treatment of distributive and penal justice to be inadequately argued. Why should we think character is any more relevant in the context of penal justice than it is in the context of distributive justice? It seems quite reasonable to assume that character influences actions relevant to distributive justice just as much as it influences actions relevant to penal justice. But if the distributive justice system should not sanction (e.g.) people whose characters are such that they lack industry, and therefore contribute less to the economy, then why should the penal justice system sanction people with character traits that lead them to crime? Rawls' theory would be simpler and better without this asymmetry.¹⁶ It can be eliminated by supposing that the principles of punishment are to be chosen in the original position along with the rest of the basic principles of society.

As we develop this idea, it is important to keep in mind that the principles of punishment chosen in the original position are only just if they are implemented along with the rest of the basic social principles that would be chosen in the original position. This has important implications. It seems safe to assume that crime rates in a society are influenced not just by penal institutions, but also by non-penal social structures such as distributions of political power, education, wealth, etc. It seems equally safe to assume that crime rates are raised by the inegalitarian distributions of these things in contemporary societies. This means that, in the dramatically more egalitarian societies in which we might justly implement the principles of punishment chosen in the original position, there would be less upward pressure on crime rates from non-penal social structures.

What should our goal be in applying original position deliberation to punishment? To begin, we have to recognize that there are unavoidable disanalogies between distributive justice and penal justice. Rawls thinks that the veil of ignorance should make us worry most about

ending up among the worst-off, and we should therefore choose principles to make that position the best it can be. But it is much harder to apply this reasoning in the context of punishment than in the context of distributive justice. This is because there is only one sort of candidate for the worst-off position when we are talking about distributive justice—the poorest—and because it is easy to think of practices that would dramatically improve things for people in this position. If we chose to, we could simply eliminate this position. That is, we could make everyone equally wealthy, by instituting steeply progressive tax rates and transferring the revenue to the less wealthy.¹⁷

Things look quite different when we turn from distributive justice to penal justice. In the case of penal justice, there are two sorts of candidates for the worst-off position—victims of crime, and the people punished—and these two positions compete in the penal justice system. That is, given human nature, it seems plausible to assume that if we change our principles of punishment to make things better for victims, we tend to make things worse for the people punished, and if we change our principles of punishment to make things better for the people punished, we tend to make things worse for the victims. We might strive to eliminate the position of victim with a police state so thorough in its control that it in effect made a prison of the world, but even this probably could not prevent all crimes, and it would in effect punish everyone. Original position deliberators would not consent to universal punishment. We could eliminate the position of the punished by ceasing to punish people altogether, but then crime rates would presumably explode, and original position deliberators would presumably consent to some form of punishment to prevent this, both to keep offenders off the street and to provide special and general deterrence (i.e. to create incentives for offenders not to reoffend, and for others not to become offenders).

So far as I can tell, there is nothing about original position deliberation that would prevent us from considering deterrence to be a legitimate goal of punishment.¹⁸ If my reasons for consenting to some form of punishment in the original position include deterrence, then the rationale for punishment in the original position is at least partly consequentialist. This means that the personhood-based justification shares something with utilitarianism. But this does not imply that the personhood-based justification a species of consequentialist justification. The premise of the personhood-based justification is that whatever we would consent to in the original position is just, not that punishment ought to be aimed at achieving any particular outcome. I take this to be a deontological premise. Whatever consequentialism appears in this justification is the result of working out the implications of this deontological premise. That is, if this line of thought is sound, it would be rational to weigh some consequentialist considerations in consenting to principles of punishment, but it is legitimate for these considerations to play a role in justifying punishment only because they emerge from our rational consent. The consequentialist considerations have no independent value. Whatever value they have derives from the value of rational consent.

It is important to emphasize that endorsing deterrence implies using criminals as *means*, but it does *not* imply using them as *mere means*. This point is crucial in defending the justification of punishment I am presenting here against the "mere means" objection. Criminals can be used as means to the end of deterrence without being used as mere means to this end if it is rational for them to consent to being used in this way. (I take this distinction between the use of others as means, and the use of them as mere means, to be a standard feature of Kantian ethics.¹⁹) In the terms of the justification of punishment I am presenting here, this claim translates as the claim that criminals can be used as means to the end of deterrence without being

used as mere means to this end if, in the original position, they would choose an institution of punishment that included deterrence.

If we have got things right so far, then original position deliberators would have to acknowledge that as long as human beings remain the imperfect creatures they currently are, our society must contain both victims and punished people, with competing interests. How would they weigh these competing interests? Presumably the point of original position deliberation is to get us to be fair to competing parties. But what would it mean to treat victims and people punished fairly? We are unlikely to find an analogue of the difference principle: it is hard to see how allowing either group to be worse off than the other could make the worse-off group better off than it would otherwise have been. Does that mean that we should set up the system so that victims and people punished suffer equally? I do not think that we should leap to this conclusion. Equalizing wealth in a society in which everyone's basic needs are met is a much more morally attractive proposition than equalizing suffering in a tragic subpopulation. We can get a better account of what fairness would consist in here by working through the procedure of original position deliberation. From this perspective, a principle is fair to competing parties if I would choose it under the assumption that I was just as likely to be harmed by it as I was to benefit. So the principles of punishment are fair if I would choose them under the assumption that I am just as likely to be the person punished as I am to be a potential victim.

I identify potential victims as the relevant beneficiaries, rather than actual victims, because potential victims arguably have more to gain from punishment than actual victims. Actual victims have already suffered the harm we would hope to avoid in the original position. For example, if the only positions I considered were those of actual victims of serious violence and punished people, and I assumed I was equally likely to end up in either position, then I might

reason that since the harm has already been done, I would gain little from punishment if I turn out to be the victim, and I would have a lot to lose if I turn out to be the punished person. I might conclude that I am better off in not endorsing any institution of punishment at all. Further, if we identify the relevant beneficiaries as potential victims, then we do not leave out anyone living who can benefit from punishment, since everyone who has been victimized can potentially be re-victimized except victims of fatal violence.²⁰

There are two further points which need emphasis. First, original position deliberation requires us to focus on what it would be like to be particular individuals in these situations—that is, a particular potential victim, or a particular punished person. We cannot focus on the welfare of victims and people punished as aggregates. The perspective of the individual person is an indispensable feature of original position deliberation, and is widely seen as part of what makes original position deliberation a deontological approach to ethics. Consequentialist approaches typically focus on aggregate well being rather than how things look from the perspective of particular individuals.

Second, in claiming that I must assume that I am equally likely to be a potential victim and a punished person, I want to rule out the idea that I can avoid becoming liable to punishment by avoiding becoming a criminal. In making this claim, I diverge from other writers who have discussed original position deliberation about punishment, including Jeffrie Murphy (1973), James Sterba (1977), and Michael Clark (2004). These writers hold that original position deliberators should assume that they will be able to avoid committing crimes, and will therefore be able to avoid punishment so long as there is not a failure of due process that causes them to be erroneously punished. Let us refer to this assumption as the "avoidability of crime assumption". I think this assumption is a mistake.

If we assume that we can control our liability to punishment by avoiding crime, our fear that we will suffer punishment ourselves will diminish, and it will be natural for us to choose more severe punishments than we would otherwise have accepted. Bear in mind that one of the purposes of original position deliberation is to design an epistemic standpoint in which the choices which come naturally to us follow the lines of justice. If we design original position deliberation in such a way as to allow the avoidability of crime assumption to diminish our fear of punishment and prompt us to endorse more severe punishment than we would otherwise accept, then we are taking the position that the avoidability of crime justifies punishment of an increased severity. But this amounts to allowing action-based desert claims to play a role in original position deliberation, and thereby opening the door to retributivism, even if it is expressed in different terminology. Sterba and Murphy acknowledge this. They think of original position deliberation about punishment as a kind of retributive justification of punishment. Clark thinks accepting the avoidability of crime assumption entails at most a negative retributivism,²¹ but for free will skeptics, this is still too much retributivism.

I think the avoidability of crime assumption is a mistake not just for free will skeptics, but also for interpreters of the original position more generally. As I explained earlier, I think that part of the point of original position deliberation is to prompt us to assume that we do not have the sort of control over our roles in society which is necessary for action-based desert. I think this view is necessary to make sense of the way original position deliberation works in the context of distributive justice. It seems essential to assume that we cannot control whether we end up among the poorest if we are to be sufficiently motivated to choose the difference principle. If we assume that we can control whether or not we end up among the poorest (for example, by working harder and earning more money) then it would be rational (at least in a self-

interested sense) to disregard the poorest entirely in choosing the basic principles of society. We would have to accept another peculiar asymmetry to assume that we can control whether we commit crimes and also that we cannot control whether we end up among the very poorest.

4. Choosing the Principles of Punishment

What principles of punishment would I choose if I had to assume that I was just as likely to be harmed by punishment as I was to benefit from it? It is crucial to recognize that the tradeoff we are talking about involves making things worse for the punished person in a tangible way in order to make things better for the potential victim in a less tangible way. That is, punishment makes the life of the punished person worse, but a reduction in someone's odds of becoming a victim of crime cannot be expected to improve his life in a correlative way. For example, any sort of imprisonment restricts the activities of prisoners in ways that can be expected to diminish their experienced quality of life, but a reduction in someone's odds of becoming a victim of crime cannot be expected to improve his experienced quality of life, so long as he is not the sort of person who worries obsessively about his odds of becoming a victim. If I could predict with certainty that instituting some particular practice of punishment would make the difference between my remaining a potential victim and my becoming an actual victim, then the benefit to me might be just as tangible as the harm to the person punished. But I cannot know this in the original position.

It may be objected that if we look at society as a whole, the benefit of punishment is just as tangible as the harm it imposes: even if we cannot know whether particular individuals will be spared victimization by an institution of punishment, we can know that there will be an overall reduction in victimization, and when we consider the overall reduction in victimization as an aggregate, it is a very tangible benefit. But this is not relevant for original position deliberation

about punishment. As mentioned earlier, one of the functions of original position deliberation is to make us think about social outcomes one person at a time. The aggregate reduction in victimization is not something that happens to a person—it is an abstraction which is a function of many people. The fact that original position deliberation disregards aggregate harm reduction is part of what makes the personhood-based justification of punishment a deontological justification. That is, it helps safeguard the personhood-based justification of punishment against the instrumentalization of criminals to which utilitarianism resorts in its unconstrained pursuit of harm reduction.

In light of these considerations, the key question we must ask is the following. How much tangible harm am I willing to impose on the person punished for the sake of bringing a less tangible benefit to the potential victim, assuming that I am just as likely to be the former as I am to be the latter?

The answer to this question may be quite different when we are thinking about different kinds of crimes. We would be unwilling to risk imprisonment to protect ourselves against theft. (This is especially true in a society that ensured that everyone had their basic economic needs met.) We would choose other forms of punishment for thieves. We would probably want thieves to compensate their victims (in order to restore the previous distribution of property) and we would probably force them to pay punitive fines to create a deterrent. We would also want to establish public institutions to compensate victims of property crimes in cases where the perpetrators cannot be identified.

We would take a different attitude toward violent crime, however. We can typically be fully compensated for crimes against our property, but this is rarely the case when it comes to violence. This is obviously true for murder, but even assaults often leave permanent bodily and

psychological injuries. We would be willing to risk imprisonment to protect ourselves against crimes of violence, so long as the conditions of imprisonment were humane. Imprisonment would prevent violent offenders from repeating their crimes, and it would provide a substantial deterrent even if prison conditions were humane. (It seems safe to assume that everyone would prefer not to have their actions restricted in the way that prison restricts action, even under humane conditions. Keep in mind that in a society in which everyone's basic needs were met, nobody would seek out imprisonment in order to get three square meals a day and a roof over his head.)

If this line of argument is sound, then we have got a significant result. That is, we can use the Rawlsian social contract to demonstrate that personhood-based desert justifies a substantive institution of punishment by humane imprisonment. Now let us consider whether this justification provides acceptable limitations on punishment.

As mentioned in the introduction, one element of the mere means objection against utilitarianism which many regard as particularly damaging is the point that utilitarianism sometimes recommends quite disturbing forms of instrumentalization. Some of these are: (i) punishments of unlimited severity; (ii) violating or weakening due process; and (iii) framing and punishing non-criminals. The idea is that if these practices turn out to give us the lowest ratio of suffering-imposed-to-suffering-prevented, then utilitarianism must acknowledge them to be the best ways to punish. It seems reasonable to hold that no justification of punishment can mount a satisfactory defense against the mere means objection unless it can rule out these practices, and this gives us three useful criteria for testing the personhood-based justification.

Let us consider limitations on the severity of punishment first. For example, now that we have justified imprisonment under humane conditions, can we go on to justify imprisonment

under harsh conditions in order to strengthen deterrence? I do not think so. Since I must make a tradeoff between a tangible harm and a less tangible benefit, and since I must assume I am just as likely to be the person harmed as I am to be the person benefited, I will not be willing to risk the kind of suffering involved in imprisonment under harsh conditions. The same would even more obviously be true of the death penalty and torture. This shows that the personhood-based justification of punishment satisfies the first criterion.

Rejecting the avoidability of crime assumption is crucial in establishing this limit. If we assume that we can avoid crime, and thereby avoid liability to punishment, it is rational in the original position to countenance very severe punishments in order to strengthen deterrence (at least if we have strong institutions of due process in place). As noted above, other philosophers who have discussed original position deliberation about punishment (such as Murphy, Sterba, and Clark) have endorsed this assumption, and given this, it becomes much more difficult for them to explain limitations on the severity of punishment.

Now let us consider the second criterion. Can the personhood-based justification explain why we should not weaken or violate practices of due process when the utilitarian calculus shows that doing so would reduce overall suffering? When we turn to this issue, the personhood-based justification's rejection of the avoidability of crime assumption may appear to be more of a liability than an advantage. In Clark's account of original position deliberation about punishment (Clark 2004), the avoidability of crime assumption plays a key role in explaining why deliberators would insist on a very strong institution of due process. That is, if we assume that we can avoid committing crimes, then it makes sense to insist on a very strong institution of due process, because this will allow us to be confident that by refraining from crime we can escape punishment. If we reject the avoidability of crime assumption, we must

find different reasons for insisting on a strong institution of due process. But the procedure of original position deliberation can readily provide us with these reasons.

As argued before, the first step in original position deliberation about principles of penal justice is to identify the people harmed and the people benefited. Previously, the focus has been on selecting the principles of punishment, that is, on determining how we should treat people who have been selected for punishment. But now, we are selecting the principles of due process: in other words, we are determining how we should select people for punishment. So, while our previous competitors were the people punished versus potential victims, the relevant competitors now are the accused versus potential victims. (The punished have little to gain from a strong institution of due process, since they have already come out on the losing end of the process, so they are not relevant competitors. The punished may of course be accused again later, but then they can be counted within the category of the accused.)

The next step in original position deliberation about principles of penal justice is to assume that I am just as likely to be harmed by them as I am to be benefited. It is clear that an individual accused of a crime (whether correctly or incorrectly) has more to lose from a weakened institution of due process than an individual potential victim has to gain from it. Suppose that the criminal conviction standard were to be lowered (from "reasonable doubt" to something weaker).²² This would in some cases allow prosecutors to convict more criminals, thereby worsening things for criminals but improving things for potential victims. But it would also open the door to sloppy or politically motivated prosecutions that would result in the conviction of non-criminals, thereby worsening things for the accused without an equivalent improvement for victims. So, on the whole, it would worsen things for the accused more than it would improve things for victims. Since I must assume that I am just as likely to be among the

accused as I am to be a potential victim, I would not choose to weaken due process. This shows that the personhood-based justification satisfies the second criterion set out above.

The issue of due process illustrates once again how important it is that the personhood-based justification does not consider aggregate harm reduction in the way utilitarianism does. Saul Smilansky (1999) argues that we would in fact reduce aggregate harm if we lowered the criminal conviction standard. He concludes that utilitarians must endorse lowering the standard, and that only deontologists can justify maintaining it at its current high level. I think he is correct, and I take this to provide additional support for the claim that the personhood-based justification is a deontological justification.

Now let us turn to the issue of framing and punishing non-criminals. Suppose that we could strengthen general deterrence by occasionally framing and punishing celebrities, given all the publicity involved, and suppose that the suffering of the celebrities would be far outweighed by the suffering prevented through crime control. Retributivists can appeal to action-based desert in order to explain why framed celebrities do not deserve punishment, but if we can only appeal to personhood-based desert, we must solve this problem differently.

This issue involves principles of punishment as well as principles of due process, so it is hard to precisely characterize the harmed and benefited parties. It is also puzzling, because it can be deeply perplexing to try to find good reasons to prefer punishing real criminals over non-criminals if we are skeptical about action-based desert. But there is a widespread intuition that it is in fact better to punish real criminals which persists in the face of such skepticism. I think this is because the intuition that it is wrong to punish non-criminals really has two roots. One root involves action-based desert, and free will skepticism pulls this root loose. The other root

involves the disturbing deception required for the framing and punishing of non-criminals to amplify deterrence, and this room remains firmly grounded if we accept free will skepticism.

Framing and punishing non-criminals could never lead to heightened deterrence unless the true goals of criminal law were completely different than its publicly professed goals, since heightened deterrence could only be achieved if almost everyone was deceived into believing that framed non-criminals were real criminals. In Kantian terms, an institution which aims to deter by way of penalizing anyone other than real criminals can only succeed through a systematic and global deception of the public which contradicts itself. That is, to choose a principle of punishment that allowed punishment of a framed celebrity instead of a real criminal for the sake of general deterrence, I would also have to choose that the overwhelming majority of the population be deceived about the fact that this principle was in effect. The deception would be necessary because if word got out that scapegoats were sometimes punished instead of real criminals, then the extra deterrent force which authorities had hoped to achieve with the framing would be destroyed. Since I could not assume that I would not be among the deceived, I would in effect be volunteering to be deceived about the principle I had chosen. In other words, I would be volunteering to be a mere means to the end of amplifying deterrence.

Consenting while one is not deceived about anything to be deceived at a later point in time does not destroy the rationality of consent in quite the same way as consenting while one is deceived. It is probably the case that we can rationally consent to being deceived in some particular cases. For example, suppose that I am pathologically afraid of taking sedatives, but I want to be sedated, because it is only if I am sedated that I will be calm enough to undergo a necessary medical procedure. In such a case I might consent to being deceived into taking a sedative—I might ask someone to lace some food item with a sedative without my knowledge.

Surely in so doing I do not make myself into a mere means in any problematic way. A life which is globally structured by rational consent may have local regions of deception. But I do not think it makes sense to suppose that a life which is globally structured by rational consent could also be globally structured by deception in the way that would be necessary for the punishment of framed celebrities to be effective. To choose to be deceived about the basic principles of one's society would be to give up on the ideal of rational consent in a way that would amount to making oneself into a mere tool. Consent to global deception undermines one's status as a rational agent in a way that parallels consent to slavery, and should be seen as self-contradictory for parallel reasons. So we could not rationally consent to an institution of punishment that punished anyone but real criminals. If this is right, then the personhood-based justification satisfies the third criterion set out above.

I am not claiming that satisfying these criteria is sufficient to show that the personhood-based justification has an ironclad defense against the mere means objection. But it seems clear that this approach can mount a strong defense. Since it can also justify a substantial institution of humane imprisonment, this seems to me to be progress well worth making.

Conclusion

If this personhood-based justification of punishment is sound, then free will skeptics have a solid alternative to utilitarianism and retributivism, and they can reply to the mere means objection. But this approach to justifying punishment should be of interest to a wider audience than just free will skeptics. Free will skeptics hold that we lack free will, and free will believers hold that we have free will. But there are many philosophers who are not willing to confidently adopt either of these positions. We might think of them as agnostics about free will. Agnostics might hold that our grounds for believing in free will are strong enough when it comes to

justifying praise and working through "ought implies can" issues, but are not strong enough to justify retribution. People with this view should be interested in a justification of punishment that has a defense against the mere means objection but does not rely on the notion of action-based desert.

Notes

Acknowledgements: Thanks to Saul Smilansky, Derk Pereboom, Joe Keim Campbell, Jeff Murphy, Erin Kelly, Kadri Vihvelin, Chad Flanders, Zac Cogley, and the editor and reviewer at *Philosophical Studies* for helpful comments on this paper or the ideas it presents. This research was supported in part by a summer grant from the Research Center for the Humanities and Social Sciences at William Paterson University of New Jersey.

¹ Some writers use these terms differently, but this is how I will use them here. (In particular, some writers refer to what I am calling retributivism as "pure retributivism").

² There are some senses of "moral responsibility" on which the truth of the claim that some human beings are morally responsible is uncontroversially compatible with the truth of determinism. Philosophers who endorse free will skepticism of the kind I address in this paper have no reason to doubt that we are morally responsible in these senses, but these senses are not sufficient to justify retribution. For example, Schlick (1939) and Smart (1961) hold that someone is morally responsible just in case praising, blaming, rewarding, or punishing him for his actions has good effects on his future actions or the future actions of others. Determinism would pose no problem for this consequentialist sense of moral responsibility, but it is not the sort needed to justify retribution—it is in fact fundamentally opposed to retributivism (see the main text below). The sense of moral responsibility which Scanlon explains in terms of "answerability", which has to do with explaining one's actions in terms of reasons one endorses, can be similarly untroubled by determinism (see e.g. Scanlon 1998). This sort of moral responsibility may be necessary for retributive justification, but it is not sufficient. The sort of moral responsibility that is arguably incompatible with determinism, which is at issue in this paper, is necessary too. (Thanks to Derk Pereboom for suggesting these points.)

³ For more on these problems, see e.g. Smilansky 1999; Primoratz 1999, chs. 2-3; Ten 1987; and Hart 1968, chs. 1-2. Some have argued that these are not really problems for utilitarianism—that if these punishments truly reduce overall suffering most, then they are just. (See e.g. Smart 1973.) I think this is a mistake.

⁴ Kant raises this objection in *The Metaphysics of Morals* (1996, p. 105, 6:331 by Akademie pagination).

⁵ I discuss Kant's theory of free will in more detail in Vilhauer 2004 and 2010.

⁶ See Clark 2004 for a helpful discussion of Kant's retributivism and resources in Kant's ethics for a non-retributive approach to punishment. As I will discuss later, however, what Clark means by "non-retributive" is still too retributive for free will skeptics.

⁷ Kant himself suggests an account along these lines at some points, but he also suggests a hypothetical consent account (discussed in the main text just below). See Scheid 1983 for a helpful discussion of both these threads in Kant's remarks on punishment.

⁸ I do not mean to claim that facts about hypothetical consent have no basis in the actual world. Presumably facts about what agents would consent to if they thought rationally about things have some basis in what agents are actually like.

⁹ Some of the remarks in this section also appear in Vilhauer 2009a. In that paper, I discuss free will skepticism and personhood-based desert, but not in the context of punishment. I repeat these remarks here in order to show how the ideas in that paper can be applied to punishment. Also see note 11 below.

¹⁰ Examples include Rachels (1978, p. 157) and Sadurski (1985, p. 131). Smilansky holds a related position, i.e., that giving up the belief that human beings are morally responsible for their actions implies giving up all our morally significant beliefs about desert (1996, pp. 157-63).

¹¹ Fred Feldman discusses this point, but not in the context of free will skepticism. (Feldman 1995a).

¹² Vilhauer 2009a.

¹³ A few of the remarks about original position deliberation in this section also appear in Vilhauer 2009a, which does not concern punishment. Also see note 7 above. I do not want to claim that the only way to unpack the idea of personhood-based desert is in terms of original position deliberation. But it seems to be a reasonable way to proceed. For a different perspective on free will skepticism which is also broadly Rawlsian, see Kelly 2002 and 2009.

¹⁴ See e.g. section 17 of *A Theory of Justice*.

¹⁵ I think that the terminology of desert is better than the terminology of entitlement for the personhood-based claims at issue here. "Entitlement" is sometimes used for claims that are in a deep sense morally arbitrary but still legitimately enforceable in a shallow sense. (For example, some ethicists might say that a wealthy farmer could be entitled to all the food grown on his lands even if his field hands were malnourished.) But the personhood-based claims at issue here are in no sense morally arbitrary.

¹⁶ This asymmetry has been discussed recently in separate papers by Saul Smilansky (2006), Eugene Mills (2004), Jeffrey Moriarity (2003), and Samuel Scheffler (2000). Scheffler gives a qualified defense of the asymmetry. Moriarity and Mills reject it, but they think this means Rawls' critique of desert should be rejected in the case of distributive justice, rather than

extended to punishment, which is the strategy advocated here. Smilansky offers an interesting account of the asymmetry which turns on the notion that almost all of us can control our actions in the way necessary to become fully deserving of the best treatment the criminal justice system has to offer (i.e. not punishing us), but few of us can control our actions in the way necessary to become fully deserving of the best treatment the distributive justice system has to offer (i.e. placing us among the richest). This view may fit what we might call non-skeptical intuitions about the scope of our control, but it is ruled out by free will skepticism.

¹⁷ This is of course not Rawls' preferred approach—his "difference principle" implies that inequality is to be preferred insofar as it makes the poorest richer.

¹⁸ It may be objected that deterrents can play no role in a free will skeptic's justification of punishment. That is, it may be thought that if a potential murderer who lacks free will refrains from murdering, then this action was inevitable, and he could not help but act in this way, so the presence of a deterrent can do nothing to explain his action. But this is mistaken. Suppose that the potential murderer has no free will because he inhabits a deterministic world. The deterrent may nonetheless be an indispensable part of the deterministic causal explanation of why he refrains from murdering. That is, it may be that if not for the deterrent, then he would have murdered. If this is true, then in an alternative possible deterministic world which contains a psychologically indistinguishable potential murderer, and which differs from the actual world only in that it lacks the deterrent, that potential murderer does in fact murder. In other words, we must not confuse determinism with fatalism. (Thanks to Saul Smilansky for suggesting a clarification of this point.)

¹⁹ It can initially sound counterintuitive to claim that it could ever be morally unobjectionable to use someone as a means, but the notion that we can legitimately use people as means is a central idea in Kantian ethics, both in Kant's own texts and in contemporary ethics literature inspired by Kant. Consider Onora O'Neill's often-cited example about interacting with a bank teller. One uses the bank teller as a means to the end of dealing with one's money, but one does not thereby use him as a *mere* means, because it is rational for him to consent to being used in this way (O'Neill 1989, 114). Whenever we achieve an end with the help of another we use that other person as a means. If it were not possible to use others as means without using them as mere means, then it would not be possible for autonomous individuals to work together. (Thanks to Derk Pereboom for a comment that prompted emphasis of this point.)

²⁰ Original position deliberation presupposes that one will find oneself among the living when the veil of ignorance is raised, so if murder victims can benefit from the punishment of their murderers, original position deliberation is committed to the view that this interest is not relevant for choosing the basic principles of society.

²¹ Negative retributivism is the view that retributivism's only role in justifying punishment is to provide an account of what means we are permitted to take to the end of harm reduction.

²² I discuss the reasonable doubt standard in more detail in Vilhauer 2009b.

Bibliography

- Clark, M. (2004). A Non-Retributive Kantian Approach to Punishment. *Ratio*, 17(1), 12-27.
- Feinberg, J. (1963). Justice and Personal Desert. In C.J. Friedrich and J.W. Chapman (Eds.), *Nomos VI: Justice*. New York: Atherton Press.
- Feldman, F. (1996). Responsibility as a Condition for Desert. *Mind* 105(417), 165-168.
- . (1995a). Desert: Reconsideration of Some Received Wisdom. *Mind* 104(413), 63-77.
- . (1995b). Adjusting Utility for Justice: A Consequentialist Reply to the Objections from Justice. *Philosophy and Phenomenological Research* 55(3), 567-585.
- Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68(1), 5-20.
- Hart, H. L. A. (1968). *Punishment and Responsibility*. Oxford: Oxford University Press.
- Kant, I. (1996). *The Metaphysics of Morals*. Tr. M. Gregor. Cambridge: Cambridge University Press.
- Kelly, E. (2002). Doing Without Desert. *Pacific Philosophical Quarterly* 83(2), 180-205.
- . 2009. Criminal Justice Without Retribution. *Journal of Philosophy* 106(8), 440-462.
- Mills, E. (2004). Scheffler on Rawls, Justice, and Desert. *Law and Philosophy* 23(3), 261-272.
- Moriarty, J. (2003). Against the Asymmetry of Desert, *Noûs* 37(3), 518-536.
- Murphy, J. (1973). Marxism and Retribution. *Philosophy and Public Affairs*, 2(3), 217-243.
- O'Neill, O. (1989). *Constructions of Reason: Explorations of Kant's Practical Philosophy* Cambridge: Cambridge University Press.
- Pereboom, D. (2001). *Living Without Free Will*. New York: Cambridge University Press.
- Primoratz, I. (1999). *Justifying Legal Punishment* (2nd ed.). New Jersey: Humanities Press.
- Rachels, J. (1978). What People Deserve. In J. Arthur and W.H. Shaw (Eds.), *Justice and Economic Distribution*. Englewood Cliffs, NJ: Prentice-Hall.
- Rawls, J. (1999). *A Theory of Justice, Revised Edition*. Cambridge, MA: Harvard.
- Sadurski, W. (1985). *Giving Desert Its Due: Social Justice and Legal Theory*. Dordrecht: D. Reidel.
- Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard.
- Scheffler, S. (2000) Justice and Desert in Liberal Theory. *California Law Review* 88(3), 991-1000.
- Scheid, D. (1983). "Kant's Retributivism". *Ethics*, 93(2), 262-282.
- Schlick, M. (1939). *Problems of Ethics*. Tr. D. Rynin. New York: Prentice-Hall.
- Smart, J.J.C. 1961. Free-will, Praise, and Blame. *Mind*, 70(279), 291-306.

- . 1973. "An Outline of a System of Utilitarian Ethics." In J.J.C. Smart and B. Williams, *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, pp. 69-72.
- Smilansky, S. (1990). Utilitarianism and the 'Punishment' of the Innocent: The General Problem. *Analysis* 50(4), 256-261.
- . (1996). Responsibility and Desert: Defending the Connection. *Mind* 105(417), 157-63.
- . (2000). *Free Will and Illusion*. Oxford: Oxford University Press.
- . (2006). Control, Desert, and the Difference between Distributive and Retributive Justice. *Philosophical Studies* 131(3), 511-524.
- Sterba, J. (1977). Retributive Justice. *Political Theory* 5(3), 349-362.
- Ten, C. L. (1987). *Crime, Guilt and Punishment*. Oxford: Oxford University Press.
- Vilhauer, B. (2004). Can We Interpret Kant as a Compatibilist about Determinism and Moral Responsibility? *The British Journal for the History of Philosophy* 12(4), 719-730.
- . (2009a). Free Will Skepticism and Personhood as a Desert Base. *Canadian Journal of Philosophy* 39(3), 489-511.
- . (2009b). Free Will and Reasonable Doubt. *American Philosophical Quarterly* 46(2), 131-140.
- . (2010). The Scope of Responsibility in Kant's Theory of Free Will. *The British Journal for the History of Philosophy* 18(1), 45-71.