Free Will Denial, Punishment, and Original Position Deliberation

Benjamin Vilhauer
Professor, CUNY City College and Graduate Center

I defend a deontological social contract justification of punishment for free will deniers.[1] Even if nobody has free will, a criminal justice system is fair to the people it targets if we would consent to it in a version of original position deliberation (OPD) where we assumed that we would be targeted by the justice system when the veil is raised. Even if we assumed we would be convicted of a crime, we would consent to the imprisonment of violent criminals if prison conditions were better than the state of nature but deterring enough to prevent the state of nature. And even if we assumed we would be accused of a crime, we would consent to an evidentiary standard low enough to allow a conviction rate sufficient to prevent the state of nature. After explaining this justification, I contrast it with the public health quarantine justification defended by Derk Pereboom and Gregg Caruso, and argue that it provides a stronger response to the using criminals as mere means objection.

A Deontological Punishment Contract for Free Will Deniers

The deontological social contract justification is founded in broadly Kantian and Rawlsian ideas about justice. This distinguishes it from other justifications of punishment available to free will deniers, which typically have consequentialist foundations. Social contract thinking plays a crucial role in formulating the justification, but it also includes moral principles

[1] This justification modifies a justification I present in previous papers (e.g. Vilhauer 2013 and 2019). I describe how it differs below.

which are *prior* to the contract and constrain the considerations to which we may appeal in contract bargaining.

The prior principles derive from a view of free will denial according to which there is a *personhood-based* kind of desert which does not depend on free will in the way *action-based* desert does. Personhood-based desert is as metaphysically and metaethically fundamental as action-based desert, and its distinctness from action-based desert means that we deserve things just by virtue of being persons even if we lack free will. Thus free will denial does not undermine personhood-based desert.[2]

Derk Pereboom and Saul Smilansky disagree with this notion of personhood-based desert. Pereboom acknowledges different kinds of desert, but holds that free will is required for what he calls *basic* desert, and holds that desert for which free will is not required (which he thinks can be derived from consequentialist considerations) is *non-basic* (2014: 2, 126-52). Thus his view appears to be that if we do not have free will, then there is nothing that anybody *fundamentally* deserves. Smilansky holds that *non-trivial* desert claims are action-based (or "responsibility-based", in his terminology). He allows that there is a minimum baseline of well-being we deserve just because we are persons, but argues that when we fall below it, our claims to deserve restoration of that baseline must be action-based (1996: 159). He illustrates this view with an example about air pollution. Smilansky argues that one deserves clean air just because one is a person, but if one's air is polluted, one deserves compensation *only if there is nothing one has done to deserve* the pollution—that is, only if the pollution does not derive from one's own culpable negligence, for example, in burning too much garbage. Smilansky thinks the

[2] I endorse free will skepticism, not denial, but I think skeptics are in the same boat as deniers when it comes to criminal justice, since retribution is unjust if we don't know we have free will.

relationship he finds in this scenario between personhood-based desert and action-based desert generalizes, such that personhood-based desert can do no non-trivial moral work unless accompanied by action-based desert claims. But the generalization does not hold, because there are things we deserve as persons which we cannot cease to deserve no matter what we do. Any human rights which are universal and cannot be alienated or forfeited are rights we deserve to be accorded no matter how we act. It is intuitive to many, laypeople as well as philosophers, that there are such rights. Personhood understood as a desert base which is metaphysically and metaethically independent from action is their only plausible grounding.

It may be objected that it is a distortion of the concept of *desert* to talk of deserving rights—that the correct concept here is *entitlement.* There are three problems with this objection. First, laypeople as well as philosophers *do* talk of deserving rights. Second, "entitlement" is often used for things people claim not for fundamental moral reasons, but instead for merely legal reasons which are morally arbitrary or even immoral. If I have a legal contract that says I own the vegetables you have grown, I am in this sense entitled to the vegetables even if your children are hungry and mine dine at gourmet restaurants. Third, the objection misses the point. What is important for the account presented here is the idea that there are personhood-based claims which derive from fundamental moral reasons which constrain legitimate punishment, and it is a merely terminological matter whether we call them entitlement or desert claims so long as we understand that they entail deontological constraints which are independent of the metaphysics of free will and moral responsibility. I use "desert" to refer to personhood-based claims in order to emphasize the point that we should not suppose that free will denial entails the collapse of all deontological foundations and leaves us with consequentialism as our only option.

The deontological social contract justification (henceforth just called the "deontological contract justification") does not claim that people deserve to be punished just because they are persons. Instead, it claims that punishment is fair if it does not violate the deontological constraints imposed by personhood. Personhood-based desert constrains punishment, but does not set an aim for punishment. As explained below, the aim is detention and general deterrence, but this aim only has normative significance because it arises from the rational consent of the punished. Thus the justification is fundamentally deontological despite having a consequentialist element.

Since free will deniers (henceforth just called "deniers") must reject action-based desert, they must not treat targets of the criminal justice system as if they deserve to suffer for their crimes. But the independence of personhood-based desert means deniers must still treat targets in accordance with personhood-based desert. So how exactly do targets deserve to be treated just by virtue of being persons? It is at this point in the story that the justification draws on contractualism. While I think it is straightforward to prompt the intuition that there are personhood-based desert claims, I do not think this intuition brings with it an immediately clear inventory of all the personhood-based desert claims we can legitimately make. Historically, laypeople as well as philosophers have typically believed in action-based desert, so there has been no intellectual pressure to disentangle action- and personhood-based desert, and there is no tradition or consensus to draw on. Instead, I draw on modified versions of Rawlsian original position deliberation (OPD) to provide a perspective that foregrounds and illuminates what we deserve from the criminal justice system based on our personhood rather than our actions. OPD is *prima facie* a useful device for this because Rawls himself recommends it in part because it screens out undeserved inequalities (1999, 86-89). Rawls' own views on free will are not clear,

and his views on which inequalities are undeserved probably differs from those deniers ought to hold, but his view of the social contract can nonetheless be useful for deniers. Rawls himself applies OPD to distributive justice, not criminal justice. The deontological contract justification extends OPD to criminal justice.[3] The deontological contract justification is conditioned on Rawlsian distributive justice. A criminal justice system chosen in OPD could only be implemented in society in a fully justified way if the distributive justice system chosen in OPD is fully implemented in society as well.

In OPD, deliberators use "maximin" reasoning to select principles that make the minimum (worst-off) social position as good as possible.[4] A disanalogy between distributive and criminal justice complicates maximin reasoning. There is only one minimum position in distributive justice – the poorest. Criminal justice has two candidates for minima, crime victims and the punished, and they compete for the worst-off position. If punishment deters, then calibrating punishment to improve one position worsens the other position. Social and technological innovations might someday eliminate this competition. Perhaps society will someday be ruled by laws so just that legitimate civil disobedience will be impossible. Perhaps swarms of soft, sagacious nanodrones will apply persistent gentle pressure to the limbs of everyone who begins to commit a crime, overwhelming them with fatigue before they succeed,

[3] OPD justifications of punishment are found in Murphy (1973), Sterba (1977), Clark (2004), and Dolovich (2004), but they rely on retributivist ideas which deniers cannot allow. Clark claims to offer a non-retributive Kantian approach to punishment, but allows what he calls "negative retributivism" in his account of why we should punish the guilty rather than the innocent, and denial rules this out. My approach is based on a more general Rawlsian approach to social justice for free will deniers presented in Vilhauer 2013.

[4] Rawls sees the risk-aversiveness of OPD as rational under uncertainty. This has been disputed (Harsanyi 1975). However, Rawls also offers a separable defense of OPD as conforming to moral facts that are prior to contract bargaining. I endorse the latter defense and am agnostic about the former.

except in cases where law-breaking is morally permissible, for example, when drivers speed to get someone to the hospital. But if criminals and victims will compete for the foreseeable future, then deniers must apply OPD in a way that is fair in view of the fact that neither party deserves to experience the consequences of crime based on their actions. The fact that Rawls does not address such competition means his account offers no guidance about how to understand fairness here. One intuitive way of thinking about fairness in competition is in terms having an equal chance of being harmed and benefited. Thus, in previous work, I explored a model of fairness in OPD bargaining about criminal justice which assumes that we have a 50/50 chance of finding ourselves among those targeted by the criminal justice system, and among potential victims, when the veil is raised (e.g. Vilhauer 2013 and 2022).[5] I continue to think that this approach has merits. Here, however, I will consider a model of fairness in which we are *certain* to be targets of the criminal justice system (to be convicted or accused of crime) when the veil is raised. This allows us OPD deliberators to work with only one minimum position, the position of the punished.

How could it be fair to assume that we will certainly be targets when the veil is raised, given that victims suffer terribly too, and compete with targets? After all, a guiding principle in OPD is to bargain solely on the basis of self-interest. The assumption that we will certainly be targets is in effect assuming that we can think reasonably about punishment under the assumption that only targets' self-interests matter, and victims' self-interests don't matter at all.

[5] Rawls himself rules out probabilistic reasoning about social outcomes behind the veil, but his target is probabilistic thinking based on assumptions about empirical distributions of outcomes in society. If probabilistic reasoning is needed in the deontological social contract justification, it must be derived from *a priori* moral ideas about how to model fairness under competition.

Two points about Rawls' conception of self-interest are important for evaluating this objection. First, OPD uses self-interest as a heuristic device—it does not claim that the only moral interests are self-interests (1999: 128). OPD restricts interests to the self-interests of people in the minimum position to ensure that the social principles it selects are fair to people in the minimum position, because the social contract as a whole is fair only if it is fair to expect everyone to consent. When the veil falls, it is often *not* in our self-interest to follow the principles we have chosen behind the veil—when the veil falls, those principles present themselves as obligations which we ought to follow for moral reasons, even if reluctantly. Second, OPD restricts interests to a *moralized* and *rationalized* conception of self-interests, which Rawls explains as interests in "primary goods…answering to [our] needs as citizens as opposed to [our] preferences and desires" (1999: xiii). Primary goods include rights, liberties, opportunities, and income and wealth (1999: 79). Perhaps even malicious desires give people interests, such that serial killers' desires to kill give them malicious interests in killing, but malicious interests are not weighed in OPD. Since the deontological contract justification appeals to contractarian thinking to unpack the implications of moral commitments which are prior to the contract, the moralization of interests in OPD does not burden the justification with additional commitments.

Objectors will press the point: even if self-interest only functions as a moralized heuristic device, that doesn't explain how an original position that accommodates self-interests of targets, but not of victims with whom targets compete, can be a model of fairness. The reason why this is a plausible model of fairness is that when we decide to punish people, we intentionally impose serious harm, and intentionally imposed serious harm must meet an especially high standard of justification. It might be thought that there is a parallel imposition of harm on *victims* by the

criminal justice system when the system fails and crimes are committed. We probably have a duty to design a criminal justice system that prevents victimization, and the occurrence of crime probably implies that we are failing to fulfil this duty and in this sense *allowing* people to become victims. But the distinction between *intentionally doing* and *allowing* is important in deontological ethics. Criminals are the ones who intentionally harm victims. Law-abiding members of society do not intentionally harm victims through the justice system.

The clearest way to be confident that we are justified in intentionally harming people is to have an account that the people harmed would accept themselves if they thought rationally about it, on the basis of their own reasons and nobody else's. Many doubt that consequentialist reasons of general deterrence suffice for such an account. Why should I accept that I should be harmed merely so that others can benefit? Retribution plays a necessary role in providing such an account in most theories of punishment. If there are reasons of retributive desert, rational wrongdoers must acknowledge that they deserve harm. But deniers reject retributive desert. Reasons of self-defense are another candidate, and they are important in the Pereboom/Caruso public health quarantine justification discussed later. But reasons of self-defense become weaker and vaguer when offenders' immediate threats have been neutralized and they are under our power, and it is at this point that the hard questions about punishment arise. The deontological contract justification's reason is OPD consent. OPD should make us more confident that it meaningfully models the consent of the punished if deliberators weigh only the interests of the punished. If OPD weighs the interests of the punished *and* the interests of those benefited by punishment, then the deontological contract justification is confronted with an objection parallel to the mere means objection: why should the interests of people who would *benefit* from my punishment give *me* a reason to consent to punishment? I do not claim to have established that

the interests of those protected by punishment do not matter at all for fairness. The point is rather that we can be especially confident that we are not imposing unfair harms on the punished if we weigh only their interests in OPD. Since intentionally imposed harm must meet an especially high justificatory standard, it is worth exploring an account that makes us especially confident that the harms we are imposing are fair. In the following, I will apply OPD which assumes that we will be targets of the criminal justice system to two kinds of targets: first, people convicted of crimes, and second, people accused of crimes.

First, how would OPD deliberators think about punishment if they assume they will find themselves convicted when the veil is raised, and they will be among the punished if any are punished? Will they consent to punishment? They will begin by interrogating the assumption. As mentioned, someday technological and social innovations may prevent all crime and thereby abolish punishment. From our standpoint outside OPD, we know our society is not post-crime. But inside the original position, deliberators are ignorant about the state of social and technological development they will find when the veil falls. They can hope they will arrive post-crime, and will choose a social principle that facilitates abolition. We might worry that deliberators in fear of punishment would seize at authoritarian measures that suppress crime but also liberty. But OPD prioritizes liberty in the primary goods pursued by deliberators, so it rules out authoritarian measures. As explained, OPD is moralized in a way that deniers should endorse. Instead, deliberators would select a principle of investing as much work as possible in non-authoritarian crime reduction measures as is compatible with meeting their other basic needs. Call this the *abolition principle*, since it aims at the abolition of punishment as an ideal that may someday become actual. Outside OPD, in today's contemporary society, this principle implies an obligation for social spending on non-authoritarian crime control. It may be beyond

the resources of the deontological contract justification to specify a spending plan with precision, but it would support the development of innovative non-authoritarian social and technological crime-control strategies, and pour money into already-available non-authoritarian crime reduction measures: more jobs, education, public services, and voluntary therapy for people at risk of committing crimes. This would add a supplement of social resources for people at risk of committing crimes onto the share allocated to them by Rawlsian distributive justice.

However, deliberators would acknowledge that they may not arrive post-crime. They know enough about human psychology (Rawls 1999: 119) to know that a society with uncontrolled crime is likely to have rising crime rates, and that rising crime rates eventually lead to a collapse into the state of nature and its war of all against all. Would they consent to some form of punishment to control crime? Consenting would be in their interest if punishment were both preferable to the state of nature and also necessary to prevent the state of nature. Let us assume that the penalties they consider for societies which are not post-crime are the ones available today, such as execution, imprisonment, and assorted lesser penalties like fines.

Deliberators would not prefer death to the state of nature, so they would not consent to a punishment system that includes execution as a penalty. Deliberators' knowledge of psychology would tell them that non-violent crime is much less threatening to social order than violent crime, and they would be reluctant to consent to imprisonment as a penalty for non-violent crime. But non-violent crimes do erode social order if unchecked. Lesser penalties like fines proportioned to wealth would provide significant checks on non-violent crime. Since fines are clearly preferable to the state of nature, it is reasonable to assume that deliberators would consent to them to control non-violent crime.

Deliberators' knowledge of psychology would also tell them that fines would not suffice to control violent crime enough to prevent the state of nature. Deliberators would therefore endorse imprisonment as a penalty for violent crime if imprisonment could offer prison conditions preferable to the state of nature while still preventing the state of nature. Deliberators would want prison conditions better than those in the contemporary U.S. prisons, because they are probably not better than the state of nature. Deliberators would want humane and secure conditions that included education, meaningful work, voluntary therapy, excellent healthcare, regular visits from friends and loved ones, continuous parole review to determine whether prisoners could be released without undue risk of repeated violence, and ongoing post-release support to help people avoid new violence.

Deliberators would want prisons' main goal to be the detention of violent offenders, not general deterrence. But they could not ignore general deterrence altogether. Saul Smilansky has argued that deniers must acknowledge that it is unjust to imprison people who do not deserve imprisonment based on their actions, and must therefore acknowledge a duty to compensate them heavily during their imprisonment (Smilansky 2011). He argues that this compensation would make prison conditions so pleasant that they would become an incentive to commit crime. While I do not agree with the entirety of this argument, I think it shows that deniers cannot simply assume that prison conditions inevitably deter, and that deniers need a moral reason to calibrate prison conditions to provide a deterrent sufficient to prevent imprisonment from becoming an incentive to commit crime. The reason offered by the deontological contract justification is OPD consent. OPD deliberators' knowledge of human psychology would inform them that excessively pleasant prison conditions would accelerate a collapse into the state of nature rather than preventing it. They would therefore want a principle that calibrates prison conditions at a level of

deterrence sufficient to avoid creating an incentive for crime, but no higher. To deter, conditions would only have to be unpleasant relative to life outside prison. Under the social conditions required by the deontological contract justification, life outside prison would be much better for people at risk of offending than it is today, for two reasons mentioned earlier. First, OPD criminal justice is conditioned on OPD distributive justice, so it entails that punishment could only really be justified when Rawlsian distributive justice is achieved, and all the incentives for crime which derive from distributive injustice were absent. Second, the abolition principle supplements the social resources allocated to people likely to offend, funding jobs, education, public services, and voluntary therapy. These social conditions make life better for likely offenders outside prison, so it can get better inside while still deterring. It seems quite plausible that, under these conditions, sufficient general deterrence could be maintained despite making prison conditions much better than the state of nature.

General deterrence inevitably uses the punished as *means*, but the consent of the punished implies that they are not used as *mere* means. Further, the punished and the people protected by punishment use each other as means *reciprocally*. Both parties seek a life better than the state of nature. The protected pursue this end by using the punished as means to general deterrence. The punished pursue this end by using the protected as means to generate the social resources necessary to provide the best prison conditions compatible with general deterrence and sufficient to prevent the state of nature. Thus there is a sense in which the competition between the punished and protected is not fundamental after all.

Next let us consider due process. Here, deliberators must make a different assumption than the one they made earlier: they must assume they have been accused of a crime but not yet convicted. Will they consent to a standard of proof for conviction that makes conviction

possible? They would prefer not to be convicted, so it might seem that they would not. But they will understand the need for punishment to prevent the state of nature in the same way the deliberators considered earlier do. So it would be irrational for them to choose a standard of proof that is impossible to satisfy. They would choose a principle that sets the standard low enough to allow a conviction rate sufficient to prevent the state of nature, but no lower. The *proof beyond reasonable doubt* standard, as it is actually applied in the many countries that endorse it, allows enough convictions to prevent the state of nature. So deliberators' principle would not allow a standard lower than reasonable doubt. Would their principle demand a higher standard than reasonable doubt? Some interpretations of the reasonable doubt standard make it extremely high. Justice Brennan, in *In re Winship*, says it demands "utmost certainty" (1970: 364). Laurence Tribe says it requires "as close an approximation to certainty as seems humanly attainable in the circumstances" (1971: 1374). On Alex Stein's interpretation, "[a] legal system may justifiably convict a person *only if it did its best in protecting that person from the risk of erroneous conviction* and if it does not provide better protection to other individuals" (2005: 175). These remarks suggest that the reasonable doubt standard demands a probability of guilt close to 100%. It is hard to see how a higher standard could allow enough convictions to prevent the state of nature. So deliberators' principle would not entail a higher standard.

The Quarantine Justification and the Mere Means Objection[6]

The mere means objection is an important objection to consequentialist justifications of general deterrence. If our rationale for punishing criminals is just that it produces the good

[6] This section draws on remarks in Vilhauer 2022. I would be happy to remove it if editors want a shorter paper.

consequence of general deterrence, we are using criminals as mere means to social ends. The deontological contract justification has a strong response to this objection. The Kantian test for whether we treat someone as a mere means is whether they would rationally consent to the treatment. The argument that we would rationally consent to punishment even if we assumed we would be convicted forms the core of the deontological contract justification. This response is stronger than the one offered by the public-health quarantine justification, which is defended by Derk Pereboom and Gregg Caruso in a number of joint papers, as well as independent papers and books (Pereboom 2001, 2014, 2021; Caruso and Pereboom 2020; Caruso 2021). They hold that the right to self-defense justifies quarantining carriers of dangerous diseases despite the fact that they do not deserve to be sick. They draw an analogy between quarantining the sick and imprisoning the violent, and argue that we have as much right to imprison the violent as we do to quarantine the sick, even if the violent do not deserve imprisonment. As I understand their overall moral theory, it is fundamentally consequentialist. Smilansky agrees (Smilansky 2019). However, they hold that the quarantine justification is *not* consequentialist because they hold (A) that the right to self-defense need not be understood consequentially, and (B) that their justification offers sufficient protection of criminals' rights not to be treated as mere means. (A) is correct, though positing non-consequentialist rights within a fundamentally consequentialist theory can be criticized on grounds of parsimony. (B) is problematic because their account of the right not to be used as a mere means is puzzling.

Pereboom's position on the right not to be used as a mere means has evolved. In *Free Will, Agency, and Meaning in Life,* he holds that people who I "harm in self-defense" are "being used merely as a means," and though this is a moral concern, it is "outweighed by the right to harm in self-defense," so long as "the harm inflicted is the minimal amount reasonably required" (2014: 167). But in non-consequentialist moral theories, the right not to be used as a mere means is

typically understood to be *absolute*, not as a right that can be overridden. From the Kantian perspective, using people as means is *never* permissible unless they would rationally consent, in which case they are *not* treated as mere means. The deontological contract justification demonstrates this with respect to punishment.

More recently, Pereboom and Caruso hold that using people as means without their consent only requires a special justification if we use them *manipulatively* toward ends other than self-defense, such as general deterrence (Caruso 2021; Pereboom 2021; see also Shaw 2019). They claim that the quarantine analogy yields what they call "free general deterrence," that is, general deterrence we from which we can benefit *without* taking on any obligation to provide a moral reason for using people as means. They think almost nobody wants to be quarantined, so quarantine inevitably produces deterrence as a *side effect*, and the same is true for imprisonment. However, as John Lemos (2016) and I (2019) have argued in different ways, quarantine does *not* inevitably deter.[7] The COVID-19 era has shown that many of us are not upset about being required to stay at home. Many still refuse to return to jobs that require them to leave the comfortable confines of home and hearth. It is not hard to imagine that the next time a pandemic strikes, many will not complain about quarantine, especially if the authorities once again send checks that eliminate income anxiety. The argument from Smilansky sketched earlier highlights the fact that when the authorities constrain people's freedoms, they have an obligation to compensate them. So the authorities have an obligation to send people checks when quarantine is imposed. But problems would arise if checks got too big. Excessive payments would give people at low risk of serious illness (because they are youthful or hardy) an incentive to intentionally expose themselves to pathogens in order to get quarantined and paid. This would make

[7] Both Lemos and I draw on arguments from Smilansky (2011).

quarantine ineffective. The authorities would need to ensure that checks were big enough, but not too big. This might seem to raise no moral concerns. Some might see any checks at all as mere largesse, a matter of supererogatory generosity, and might therefore suppose that the authorities could limit the checks as they wished without providing a moral reason for the limit. But this is a mistake, because the checks are an obligatory response to the authorities' constraints on liberties, and the authorities must therefore have a moral reason for setting limits on their response. Said differently, quarantine practices must be calibrated to make life in quarantine as pleasant as possible while still making it *unpleasant* enough to *deter* people from intentional exposure. Calibrating quarantine in this way intentionally makes quarantine sufficiently unpleasant. We need a moral reason to do this, and the moral reason we need is quite precisely a moral reason for general deterrence. This means that general deterrence does not come as a "free" byproduct – we cannot justify effective quarantine without justifying general deterrence.

If the analogy Pereboom and Caruso draw between quarantine and punishment holds up, then justifying effective punishment also requires justifying general deterrence. Arguments earlier in this paper show this to be true. But here is an example to illustrate the point. Suppose I am a prison warden, and my goal is merely to detain violent offenders in comfortable conditions. Suppose I believe I am not entitled to aim at general deterrence, since I think calibrating prison conditions for general deterrence would nonconsensually and impermissibly use the imprisoned. Now suppose I discover that conditions are producing general deterrence as a side effect. Perhaps general deterrence comes for free until I discover this, since I did not intend to create general deterrence. But upon discovery, it is no longer free: my belief about the impermissibility of calibrating conditions for general deterrence gives me a reason to *improve* conditions. If I wish to preserve general deterrence, I need a justification of general deterrence. I could appeal to

the deontological contract justification to show why people would rationally consent to being used for general deterrence, or I could.appeal to consequentialism and argue that it is permissible to nonconsensually use people for general deterrence after all.

In response to considerations like this, Pereboom now endorses a view that supplements free general deterrence with a straightforwardly consequentialist argument for general deterrence (2021: 91, 102), and holds that even nonconsensual, manipulative use is consistent with people's right not to be treated as mere means (95) so long as they are not treated too severely (85). This distorts the right not to be used as a mere means. Caruso still holds that the only general deterrence which deniers should endorse is free general deterrence (2021: 312). But there is no free general deterrence, so appeals to it obscure a justificatory burden that punishment theories should be able to bear. The deontological contract justification can bear it.

Conclusion

OPD deliberation that assumes we will be targets of the criminal justice system gives us high confidence that we can justify a criminal justice system with a reasonable doubt conviction standard and prison conditions deterring enough to prevent the state of nature. But as discussed, there is room for debate whether the criminal justice system we endorse under the assumption that we will be among its targets is fair to the people the criminal justice system should protect. I have not argued that it is fair to them. Instead, I have argued that we can be especially confident that we will not treat targets unfairly if we follow the principles to which we would consent under the assumption that we will be targets. A model of fairness in which OPD deliberators weigh interests of both the targeted and protected may justify a more stringent criminal justice system, one which prevents more crime through more unpleasant prisons and a burden of proof which is easier for prosecutors to satisfy. The arguments here imply that we can have less

confidence that such a system is fair to its targets. But as far as the argument here is concerned, it *may* be fair to its targets despite our lower confidence. However, prominent critics of free will denial sometimes argue that deniers cannot justify punishment at all (Smilansky 2011), and sometimes argue that deniers must endorse practices that strike our moral intuitions as excessively punitive (Lemos 2023, especially chapters 7 and 8). So it is valuable for deniers to have a justification which offers high confidence that we can treat targets fairly despite treating them stringently enough to avoid the state of nature.

References

Caruso, Gregg D. 2021. *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*. Cambridge: Cambridge University Press.

Caruso, Gregg D., and Derk Pereboom. 2020. "A Non-Punitive Alternative to Punishment." In *The Routledge Handbook on the Philosophy and Science of Punishment*, edited by Farah Focquaert, Elizabeth Shaw, and Bruce N. Waller, 355-65. New York: Routledge.

Clark, Michael. 2004. "A Non-Retributive Kantian Approach to Punishment." *Ratio* 17, no. 1 (February): 12-27.

Dolovich, Sharon. 2004. "Legitimate Punishment in Liberal Democracy." *Buffalo Law Review* 7, no. 2 (January), 307-442.

Harsanyi, John. 1975. "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory." *American Political Science Review* 69, no. 2 (June): 594-606.

Lemos, John. 2023. *Free Will's Value.* New York: Routledge.

———. 2016. "Moral Concerns about Responsibility Denial and the Quarantine of Violent Criminals." *Law and Philosophy* 35, no. 5 (October): 461-83.

Murphy, Jeffrie G. 1973. “Marxism and Retribution.” *Philosophy and Public Affairs* 2, no. 3 (Spring): 217-43.

Pereboom, Derk. 2001. *Living without Free Will.* Cambridge: Cambridge University Press.

———. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

———. 2021. *Wrongdoing and the Moral Emotions.* Oxford: Oxford University Press.

Pereboom, Derk, and Gregg D. Caruso. 2018. “Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life.” In *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, 193-222. Oxford: Oxford University Press.

Rawls, John. 1999. *A Theory of Justice*. Rev. ed. Cambridge, MA: Harvard University Press.

Shaw, Elizabeth. 2019. “Justice without Moral Responsibility?” *Journal of Information Ethics* 28, no. 1 (Spring): 95-130.

Smilansky, Saul. 1996. “Responsibility and desert: Defending the connection.” *Mind* 105 (417), 157-163.

———. 2000. *Free Will and Illusion*. Oxford: Clarendon.

———. 2011. “Hard Determinism and Punishment: A Practical *Reductio*.” *Law and Philosophy* 30, no. 3 (May): 353-67.

———. 2019. “Free Will Skepticism and Deontological Constraints.” In *Free Will Skepticism in Law and Society: Challenging Retributive Justice*, edited by Elizabeth Shaw, Derk Pereboom, and Gregg D. Caruso, 29-42. Cambridge: Cambridge University Press.

Stein, Alex. 2005. *Foundations of Evidence Law.* Oxford: Oxford University Press. 175

Sterba, James P. 1977. “Retributive Justice.” *Political Theory* 5, no. 3 (August): 349-62.

Supreme Court of the United States. 1970. *In re* Winship, 397 U.S. 358.

Tribe, Laurence H. 1971. “Trial By Mathematics: Precision and Ritual in the Legal Process.” *Harvard Law Review* 84, no. 6 (April): 1329-1393.

Vilhauer, Benjamin. 2009. “Free Will Skepticism and Personhood as a Desert Base.” *Canadian Journal of Philosophy* 39, no. 3 (September): 489-511.

———. 2013. “Persons, Punishment, and Free Will Skepticism.” *Philosophical Studies* 162, no. 2 (January): 143-63.

———. 2022. “Free Will Skepticism and Criminals as Ends in Themselves”. In Matthew C. Altman (ed.), *The Palgrave Handbook on the Philosophy of Punishment*. Palgrave-Macmillan. pp. 535-556.