JPE

Preemptive Incapacitation, Victim's Rights, Desert, and Respect for Persons: Replies to McCormick and Donelson

GREGG D. CARUSO
SUNY Corning and Macquarie University

Keywords: free will, moral responsibility, punishment, criminal justice, retributivism, quarantine, public health, free will skepticism, incapacitation, victim's rights, desert, psychopathy, preemptive incapacitation, criminal law, rehabilitation, respect, legal punishment

I would like to begin by thanking Kelly McCormick and Raff Donelson for their thoughtful and challenging comments on my book, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (2021a). They have both given me much to think about. While Donelson generally agrees with the dual aims of my book (i.e., rejecting retributivism and replacing it with the public health-quarantine model), he parts ways with me on 'certain arguments and certain ways of framing the debate' (2022). McCormick, on the other hand, has serious concerns about the implications of my public health-quarantine model. In what follows, I will attempt to address each of their concerns and convince them that the retributive justification of legal punishment faces insurmountable difficulties and that the public health-quarantine model remains the best alternative. I begin with my replies to Kelly McCormick.

1. Replies to Kelly McCormick

McCormick focuses her comments on the implications of the public health-quarantine model itself. In particular, she raises a dilemma for the model

Contact: Gregg D. Caruso <gcaruso@corning-cc.edu>
https://orcid.org/0000-0002-9981-4922

https://doi.org/10.3998/jpe.3547

concerning whether (1) it is intended as an *ideal* model, whereby we ought to radically overhaul our approach to criminal behavior and fully embrace a shift to *prevention* (rather than sanctions) as a foundational aim, or (2) it is intended as a more *pragmatic* model, whereby we ought to go as far as we can toward revising our approach to criminal behavior while still working within the bounds of our actual criminal-justice practices as we find them. Against the former (ideal) reading, she argues that concerns about respect for the human dignity of offenders linger, despite my efforts to defuse them. Against the latter pragmatic reading, she argues that safeguarding the human dignity of offenders comes at the expense of the dignity of their victims. In the end, McCormick maintains that the public health-quarantine model is faced with a troubling tension regarding our respect for offenders and our respect for victims, and it's not obvious how this tension might be resolved.

1.1 Involuntary Mental Health Therapies and Preemptive Incapacitation

With regard to the first horn of the dilemma, McCormick focuses on two potential concerns regarding the human dignity of offenders: involuntary mental health therapies and preemptive incapacitation. With regard to the first concern—involuntary mental health therapies—she focuses on how my model would deal with psychopaths. She writes, 'It is not clear how such a model can thread the needle between achieving its preventative aims on the one hand and respecting the human dignity of psychopaths on the other' (XX). She continues on and adds:

> One obvious policy for dealing with psychopathic criminal behavior would be to incapacitate those diagnosed with psychopathy as soon as they have been deemed a significant threat to the safety of others. However, this policy would be at odds with the spirit of Caruso's public health-quarantine model. Such diagnoses could—and often do—occur at a very young age. A model that recommends incapacitating *children* or even young adults due to the threat to public safety posed by psychopathy would be sharply at odds with Caruso's repeated assertions that the public health quarantine model does not require incapacitating the *innocent*. (XX)

McCormick is correct that I oppose preemptively incapacitating individuals diagnosed with psychopathy. In fact, I generally oppose the use of preemptive incapacitation in all but the most extreme cases (see Caruso 2021a, 2021b; Pereboom and Caruso 2018). More on this point in a moment.

With respect to psychopathy, more specifically, let me first note that there is no standard test of it in children. Psychologists recognize psychopathy only as an adult disorder. Psychologists, however, may diagnose children with callous-unemotional traits, which are believed to be an early indicator of psychopathy. In the book I note:

> While a public health approach to criminal behavior should welcome any and all improvements in our current risk assessment instruments, including those provided by neuroscience, I fear that these measures can potentially be used to justify preemptive incapacitation for those who are deemed a risk to society. There is also the very real potential for stigmatization—identifying children who exhibit early psychopathic traits, for example, may be helpful in providing early interventions, but it can also stigmatize them by labeling them as potential future criminals. (2021a: 246–47)

For this reason, I make several broad suggestions on how to deal with such cases. First, I argue that on my model, preemptive incapacitation should be prohibited in all but the most extreme circumstances (see below). Second, since the potential for stigmatization in youth who exhibit callous-unemotional traits is a serious one, I recommend that we favor interventions that (1) maximize the autonomy of agents, (2) acknowledge the potential for change, and (3) focus on current antisocial behavior rather than future risk of offending (247). In fact, studies have found that not all children who exhibit callous-unemotional traits grow up to be adult psychopaths, which is important. This challenges us to find the right interventions—and there are some promising treatment approaches out there for young people.

This, however, brings me to McCormick's next concern: 'the degree to which those diagnosed with psychopathy—especially at a young age—would be given a genuine, voluntary choice to undergo such therapy on Caruso's model' (XX). According to McCormick, 'If . . . the only alternative offered by the state is a lifetime of incapacitation, then the answer seems clearly no' (XX). In response to this concern, I offer three replies. First, I reject the false dilemma between 'a lifetime of incapacitation' and forced therapy. This is because I reject preemptive incapacitation for simply being diagnosed with psychopathy. Not everyone diagnosed with psychopathy will engage in violent and criminal behavior. Furthermore, given that we are unable to assess with certainty the likelihood of future violent behavior, and given the potential for false positives, I maintain that significant weight should be given to protecting individual liberty.

Second, consent is a complicated issue, even when the sanctions for noncompliance fall far short of incapacitation. Consider, for example, schoolchildren

with attention-deficit hyperactivity disorder (ADHD). Public schools are not allowed to require any student to take ADHD medication. Schools are, however, allowed to share concerns with parents about a child's behavior or performance in school. Teachers might describe what they've seen or how a child's behavior is getting in the way of learning. They can even suggest or request that a child be evaluated for special education. If all else fails, however, the school can ask the child to leave on the grounds that they cannot behave. Does this mean that consent is undermined since the school's message to the parent could be interpreted as 'your child needs to be medicated in order to stay here'? I think it's complicated, but I do *not* think it's impermissible for a school to share concerns with parents, request a child to be evaluated, and even expel students for bad behavior. I would say the same for those diagnosed with psychopathy—or, in the case of a child, callous-unemotional traits. They should retain the right to refuse various therapies, but the state (and its representatives) likewise retains the right to protect itself and its members by imposing liberty-limiting restrictions when necessary.

Lastly, I would stress that even on the ideal interpretation of the public health-quarantine model, the goal of prevention does not override all other considerations. In fact, throughout the book, I argue that my model retains respect for persons,[1] prohibits the manipulative use of agents, and gives significant weight to concerns about liberty and autonomy. For this reason, I deny that it is a purely utilitarian or consequentialist theory, one that is willing to do anything to produce a net increase in overall safety. Instead, my model is constrained by a number of considerations, including respect for persons, the conflict-resolution principle,[2] the principle of least infringement, the principle of normality, and the prohibition of manipulative use.

Let me now turn to McCormick's concern about preemptive incapacitation. Here, she worries about a policy proposal I make regarding the use of predictive technologies. She correctly notes that I recommend making use of big data and other predictive technologies to aid in identifying, tracking, and predicting violent behavior for the purpose of designing general and local interventions but

1. For more on why there is no inconsistency between adopting free-will skepticism and maintain respect of persons, see Pereboom (2001, 2014), Vilhauer (2009, 2013a, 2013b), and Caruso (2021a). See also my sec. 2.2 below.

2. According to the *conflict-resolution principle*: When there is a significant threat to public health and safety, individual liberty can be limited but only when it is (1) in accordance with the right of self-defense and the prevention of harm to others, where (2) this right of self-defense is applied to an individual threat and is calibrated to the danger posed by that threat (not some unrelated threat), and (3) it is guided by the principle of least infringement, which holds that the least restrictive measures should be taken to protect public health and safety (Caruso 2021a: 192).

recommend *prohibiting such technology from being used for the purpose of preemptive incapacitation*. She wonders how I can justify my caveat prohibiting the use of predictive technologies for such purposes. She writes:

> I find Caruso's appeal to *innocence* in order to block concerns about preemptive incapacitation puzzling, given his commitment to free-will skepticism. As others have persuasively argued . . . once we have abandoned the notion of basic desert of blame, *everyone* is innocent, even those who have engaged in violent criminal behavior. As such, if it is permissible to incapacitate those who have engaged in such behavior, what exactly blocks our justification for preemptively incapacitating others who have not yet caused harm but for whom we have good reason to think such behavior is likely in the future? The intuitive prohibition on harming (or, in this case, incapacitating) the innocent seems fundamentally grounded in the fact that the innocent have done nothing to *deserve* this harm. But as a free-will skeptic, Caruso must exercise caution not to sneak in appeals to basic desert. (XX)

In response, I would first like to note that there are (at least) two significations of the relevant notion of *innocence* that need to be distinguished here. The first has do with the kind of innocence McCormick has in mind—the kind that is best captured in terms of *basic desert*. Here, innocent agents are not *deserving* of punishment, blame, resentment, indignation, or moral anger, and hence it would be wrong to preemptively cause them harm on the grounds that they deserve it. McCormick is correct that as a free-will skeptic I am unable to appeal to this notion of innocence, since I reject the grounds on which it is drawn—basic-desert moral responsibility.

There is, however, another standard sense of innocence that does not appeal to controversial notions of free will and basic desert. It's the sense in which an agent is innocent if they are not causally responsible for intentionally committing a crime. While causal responsibility is backward-looking, since it requires us to look back to identify the cause of some effect, it is in no way predicated on the notion of basic desert and is completely compatible with the rejection of free will. This is important since the kind of incapacitation licensed by the right of self-defense generally does not apply to those who are innocent in this latter sense. If Sally, for instance, is not causally responsible for shooting Bertrand, since it was Fred who actually did it, then it would be wrong on my account to incapacitate Sally since the right of self-defense and defense of others does not apply to those who are nonthreats. There is, then, a sense of innocence that plays an important role on my account, one that demands the state prove beyond a reasonable doubt that it was *this specific agent* who committed the crime and not

some other, otherwise we run the risk of incapacitating nonthreats, which my model prohibits.

That said, I take McCormick's concern about preemptive incapacitation to be about those innocent agents who have not yet caused harm but whom we have 'good reason' to think such behavior is likely in the future. While I take this concern seriously, I maintain that there are several good reasons for opposing the preemptive incapacitation of competent, reasons-responsive agents in practically all real-world cases (see Caruso 2021a, 2021b; Pereboom and Caruso 2018). First, I argue that the right to liberty must carry weight in this context, as should the concern for using people merely as means. Second, the risk posed by a state policy that allows for preventative detention of non-offenders needs to be taken into serious consideration. In a broad range of societies, allowing the state this option stands to result in much more harm than good, because misuse would be likely. Third, while the kinds of testing required to determine whether someone is a carrier of a communicable disease may often not be unacceptably invasive, the type of screening necessary for determining whether someone has violent criminal tendencies might well be invasive in respects that raise serious moral issues. Moreover, available psychiatric methods for discerning whether an agent is likely to be a violent criminal are not especially reliable, and detaining someone on the basis of a screening method that frequently yields false positives is seriously morally objectionable (Morse 1999; Nadelhoffer et al. 2012).

For these reasons, I propose we adopt an attitude of epistemic skepticism when it comes to judging the dangerousness of someone who has not yet committed a crime. Given the limitations of our current screening methods, their invasiveness, and the likelihood of false positives, our default position should be to respect individual liberty and prohibit the preventative detention of non-offenders. Additionally, Jean Floud and Warren Young (1981) have argued that anyone who has not yet committed a crime should be entitled to a *presumption of harmlessness*, much as a person should be entitled to a *presumption of innocence*. Just as the presumption of innocence protects the unconvicted person against punishment, so the presumption of harmlessness protects the unconvicted person against preventive detention. And not only is the presumption of harmlessness consistent with the attitude of epistemic skepticism, it is also a presumption that should be afforded all rational individuals since respect for persons and considerations of justice demand it. Persons, qua persons, are deserving of respect (see sect 2.2) and, as such, should be granted the presumption of harmlessness so as to protect them from unjustified and hasty restrictions on their liberty.

These considerations, I contend, will block preemptive incapacitation in all but the most extreme cases. But where and when should exceptions be made?

What are the 'extreme cases'? Here is one possible case: 'Imagine that someone has involuntarily been given a drug that makes it virtually certain that he will brutally murder at least one person during the one-week period he is under its influence. There is no known antidote, and because he is especially strong, mere monitoring would be ineffective' (Pereboom and Caruso 2018: 215). In such a case, the public health-quarantine model may allow for preventative detention. But this should not count as a strong objection to the view since virtually everyone would agree that it would be at least prima facie permissible to preventatively detain this individual for the week. In fact, instead of being a weakness of the model, I think the possibility of exceptions in cases like this provide the model with an additional advantage over retributivism. In this situation, retributivists would have a hard time justifying preventive measures on the grounds that the individual deserves punishment, since he has not yet done anything wrong. Yet the retributivists I have talked to about this example, as well as the many audiences I put the question to, have acknowledged that they too would detain the individual until the drug has worn off. This indicates to me that most retributivists are not pure retributivists but are willing to supplement their account with additional justifications when needed, such as the justification provided by the public health-quarantine model. But if one is going to be a *retributivist plus* (i.e., if one is going to embrace retributivism but supplement it with additional justifications, like incapacitation, when needed), then they too must address concerns about preemptive detention.

What, then, justifies the use of preemptive incapacitation in the drugged-agent case but not in most other cases? I maintain that it's not simply the level of certainty we have that the agent will brutally murder at least one person while under the drug's influence. Rather, it's the fact that the drug has undermined the agent's reasons-responsiveness, and that in such cases it is sometimes permissible to preemptively incapacitate nonreasons-responsive agents, as is currently permitted in cases of mental illness when an individual is involuntarily committed for being an imminent and serious threat of substantial harm to self or others.[3]

3. Note that free-will skeptics do not deny the importance of reasons-responsiveness. They simply deny that responsiveness to reasons is sufficient for the kind of control in action required for basic-desert moral responsibility, as some contemporary compatibilists maintain (e.g., Fischer 2007; McKenna 2012, 2020; Sartorio 2016). In fact, on the forward-looking account of moral responsibility developed by Pereboom, responsiveness to reasons is required for an agent to be subject to moral formation on the basis of presentation of reasons (see, e.g., Pereboom 2014, 2021; Caruso 2021; Caruso and Pereboom, forthcoming). On this account, in the face of wrongdoing we can request an explanation with the intent of having the agent acknowledge a disposition to act badly, and then, if they have in fact acted without excuse, aim for them to come to see that the disposition issuing in the action is best modified or eliminated. This change is produced by the recognition of moral reasons to make it. Thus, it's an agent's responsiveness to reasons, together with our

In fact, existing law already allows for preventive involuntary commitment in certain rare and specifically defined cases (e.g., the involuntary commitment of the dangerously mentally ill). The justification for such commitment lies largely in the fact that such individuals are not just dangerous but both dangerous and unable to confirm their behavior to the law (whether from a deficit of reason or of will). I see no reason why the public health-quarantine model cannot also make use of this distinction between agents who are suffering from mental illness and satisfy the legal conditions for involuntary commitment and agents who are competent and reasons-responsive. I propose that in cases when an individual lacks competency, is suffering from a serious mental illness, and poses a real and present threat of substantial harm, involuntary commitment may be justified, as long as it is also guided by the principles of least infringement, beneficence and nonmaleficence, and concern for the well-being of the individual. I contend the same is true for the drugged man since he likewise is not in control of his actions, is 'virtually certain' to kill at least one person during the next week if not incapacitated, and is unable to make rational and informed decisions during the one-week period he is under the drug's influence.

On the other hand, I contend that when agents are competent, moderately reasons-responsive, and able to make rational and informed decisions, the presumption of harmlessness and epistemic skepticism count strongly, perhaps even conclusively, in favor of allowing individuals the liberty to commit criminal acts before incapacitation can be justified. Moderately reasons-responsive agents are able to conform their behavior to the reasons they have for action

interest in their moral formation and our reconciliation with them, that explains why they are an appropriate recipient of desert-free moral protest (for more on this account, see Pereboom 2021; Caruso and Pereboom, forthcoming). Free-will skeptics also maintain that reasons-responsiveness is crucial for determining the right response to crime (Pereboom and Caruso 2018; Caruso 2021). On the public health-quarantine model, for instance, the question of whether an offender is reasons-responsive is relevant for at least two reasons. First, it is important in assessing what kind of threat an individual poses moving forward and whether incapacitation is required. An offender suffering from a serious mental illness, for instance, differs in significant ways from one who is fully reasons-responsive. These differences will be relevant to determining what minimum restrictions are required for adequate protection. Second, if the capacities of reasons-responsiveness are in place, forms of treatment that take rationality and self-governance into account are appropriate. On the other hand, those who suffer from impairments of rationality and self-governance would need to be treated differently and in ways that aim to restore these capacities when possible. Understanding the variety of causes that lead to impairment of these capacities would also be crucial to determining effective policy for recidivism reduction and rehabilitation. Hence, free-will skepticism and the public health-quarantine model acknowledges the importance of reasons-responsiveness, self-governance, and differences in degrees of autonomy. But rather than see these as relevant for assigning blameworthiness and basic-desert moral responsibility, they instead view them as important (in fact, essential) to determining the appropriate course of action moving forward.

(Fischer and Ravizza 1998). If we think that a reasons-responsive agent is going to harm another, we could reason with them or provide them with countervailing moral and/or legal considerations, but we should respect their liberty to conform with those reasons or not. And given that we have no way of determining ahead of time what reasons will ultimately move them, we should afford them the presumption of harmlessness.

I acknowledge that it is possible that this epistemic burden of proof can be overcome in certain hypothetical situations, but in most real-world cases, since we lack a crystal ball, we are nowhere near certain that a reasons-responsive agent will cause severe and substantial harm to others until they do so. Furthermore, even when we have good indication that someone poses a serious threat of harm, there are countervailing moral considerations that must be taken into account (e.g., the principle of respect for persons [which I argue is consistent with my free-will skepticism], the likelihood of false positives, and the risk posed by a state policy that allows for preventative detention of non-offenders). I therefore argue that the preemptive incapacitation of competent, reasons-responsive agents should be prohibited. And in the cases where my model would allow exceptions, like the hypothetical example of the drugged man, virtually all theorists would agree that some preemptive measures should be taken.

I would also suggest that in many real-world cases, concerns about preemptive incapacitation are not pertinent. That is because the actions and behaviors we often take to indicate that someone is a real, present, and substantial threat to others are *themselves* criminal acts, or at least could be interpreted as such, in which case incapacitation, if it were deemed justified, would no longer be preemptive. Instead, it would be a reaction to criminal offenses already performed and hence justified on the right of self-defense. For instance, if a would-be school shooter makes a video expressing their intentions and plans, and then posts it online before attempting the actual shooting, the video itself would likely be interpreted as a criminal act and itself grounds for restrictive measures. Similarly, a violent stalker who has not yet harmed their intended target may violate privacy and cyber laws before their action escalates to the point of violent behavior. If these are the signs by which we judge that an individual is a significant threat to others, then any restrictive measures we take would no longer be preemptive; they would be responsive to an actual crime.

Since most real-world grounds for thinking an individual poses a significant threat of harm will involve prior actions of this kind, the realm of cases where preemptive incapacitation is even a legitimate question may be vanishingly small. They will generally only arise when the grounds for thinking someone poses a significant and present threat come not from their actions or behavior but

from sociological and/or neuroscientific data indicating an increased risk of violence. But this is exactly why I argue in the book that such data are insufficient to overcome the presumption of harmlessness and the epistemic burden of proof. This, then, is the basis for my caveat prohibiting the use of predictive technologies for the purpose of preemptive incapacitation. Imagine, for example, that the social and neurological determinants of crime are like individual dials on a vast combination lock. Even if nineteen out of, say, twenty dials are in place, the lock will not open until the last dial (e.g., the last environmental or neurological trigger) is put in place. Of course, in real life the number of variables responsible for violent behavior is much greater. Since we are in a poor epistemic position to judge when (if ever) certain factors will trigger a violent episode, and since this will likely remain true for the foreseeable future, we should put a bright line in the sand in favor of protecting individual liberty and against incapacitation—*especially* with regard to risk assessed by brain scans or other demographic risk factors.

1.2 Rehabilitation, Reintegration, and Respect for Victims

I now turn to the second horn of McCormick's purported dilemma, the one directed at a pragmatic interpretation of my public health-quarantine model. On this reading, McCormick writes, 'Rather than aiming to *prevent* criminal behavior as much as possible, perhaps the idea is simply that we enact common-sense policies for substantially reducing criminal behavior and the need for incapacitation within the confines of many of our current practices as we find them' (XX). Against this interpretation, McCormick argues that safeguarding the human dignity of offenders comes at the expense of the dignity of their victims. She writes: 'My concern is that some of the policies that Caruso endorses as required by the public health-quarantine model are not always consistent with respect for the human dignity of victims. In fact, for many kinds of violent and sexual crimes, meeting Caruso's own conditions for respecting the dignity of offenders will be downright *inconsistent* with also respecting their victims' (XX).

In response, I would first note that I reject the purely pragmatic reading of my proposal. The public health-quarantine model should not be understood as simply recommending a set of pragmatic suggestions for reducing criminal behavior within the confines of 'our current practices as we find them'. It is a more radical proposal than that. That said, I am also not completely comfortable with McCormick's caricature of the ideal interpretation. This is because the suggestion seems to be that on the ideal view, prioritizing prevention means overriding all other considerations—including respect for

persons and concerns about autonomy. This is not what I claim. If I were forced to choose between McCormick's two interpretations, I would say that the ideal reading is closer to what I intend but with the key proviso that important moral constraints remain in place and must be weighed heavily (constraints that are consistent with the rejection of free will and basic-desert moral responsibility).

Despite, though, my rejection of the more pragmatic interpretation, it remains important that I address McCormick's more specific concern about respect for victims. McCormick provides the following all-too-realistic-and-sad example:

> Imagine a case in which the victim, Aaliyah, is a university undergraduate at a small liberal arts college. One night at a party, Aaliyah has something slipped into her drink and she is sexually assaulted by one of the hosts. Let's consider a version of this kind of case, in which Aaliyah reports the crime, agrees to testify in court, several witnesses come forward to support her, her case is rigorously prosecuted by the district attorney, and her assaulter, Brock, is ultimately found guilty. On Caruso's model, what should the state do with Brock? (XX)

McCormick is concerned that my model would not be able to justify incapacitating Brock, but I disagree. In such a situation, the defensive rights (the right of self-defense and prevention of harm to others) could easily justify incapacitating Brock since he remains a significant threat to the safety of other women. In fact, numerous studies show that sexual offenders are often repeat offenders (see, e.g., Lisak and Miller 2002; Foubert, Clark-Taylor, and Wall 2020). For instance, Foubert, Clark-Taylor, and Wall studied fifty-three college campuses and found that more than 87 percent of alcohol-involved sexual assault was committed by serial perpetrators. So, contrary to McCormick's suggestion to the opposite, the public health-quarantine model would likely recommend incapacitating Brock.[4]

McCormick, however, thinks that *even if* that's the case, serious problems remain. She gives an example where Brock is sentenced to a year of incapacitation at a cushy prison, then released and allowed to return to the same college as Aaliyah. To further complicate matters, McCormick writes: 'Aaliyah was a freshman at the time of her assault, and she and Brock share the same relatively small philosophy major. As such, it will now be impossible for Aaliyah to avoid Brock both on campus and in class while also completing her own coursework for

4. I say 'likely' because hypothetical cases like this are hard to be definitive about without all the relevant details needed to make a final judgment.

her major' (XX). McCormick justifiably worries that adopting a 'more humane' approach toward offenders like Brock will result in the further victimization of Aaliyah—since, in such a case, Aaliyah might be forced to 'change her major, if not leave the school entirely, regardless of her own preferences'.

While I fully share McCormick's concerns about the victims of sexual assault, and the need not to further victimize them, I do *not* think the problem lies in adopting my more humane approach toward criminal behavior. First, it's a mistake to think that the public health-quarantine model would simply recommend releasing Brock after a year regardless of whether or not he has been rehabilitated. If Brock remained a serious threat to society after the year, my model would recommend continued incapacitation. Second, the principle of least infringement maintains that the least restrictive measures should be taken to protect public health safety. That does not mean, however, that we could not justifiably impose other liberty-limiting restrictions on Brock after, or in lieu of, incapacitation—including prohibiting Brock from returning to the same college as Aaliyah. Such restrictions would be justified on the grounds of self-defense and prevention of harm to others and would resemble many of the other restrictions we typically place on sexual offenders. We regularly prohibit convicted child molesters, for instance, from being within a certain distance from parks and schools, and we do so not on retributive grounds but because we think it necessary to protect public safety. In similar fashion, the public health-quarantine model could appeal to the need to protect Aaliyah from further harm to justify similar restrictions on Brock.

Furthermore, it's unclear what alternative McCormick has in mind. If she thinks 'safeguarding the human dignity of offenders [comes] at the expense of the dignity of their victims', does she recommend *removing* those safeguards? I think that would be ill-advised. Additionally, it's unclear that retributivism would do a better job in this situation. That's because retributivism is concerned with giving wrongdoers their just deserts, *not* satisfying the needs and desires of victims. As the retributivist Alec Walen accurately points out, 'The view that it wrongs victims not to punish wrongdoers confuses vengeance, which is victim-centered, with retributivism, which is agent-centered: concerned with giving the wrongdoer the punishment *he* deserves' (2014). Paul Robinson (2008) has further argued that retributivists must distinguish between vengeful and deontological conceptions of deserved punishment. The former urges punishing an offender in a way that mirrors the harm or suffering he/she has caused. The problem, however, is that justifying punishment on the grounds of vengeance or the satisfaction of retributive hatred fails to take into account the blameworthiness of the offender. The deontological conception of desert, on the other hand, focuses not on the harm of the offense but on the blameworthiness of the offender, as drawn from the arguments and analysis of moral philosophy. To the

extent, then, that retributivists want to appeal to moral blameworthiness rather than vengeful desires in justifying punishment, denying victims the vengeful satisfaction they seek would not be a violation of their rights.

This brings me to my next reply: punishment inflicts harm on individuals, and the justification for such harm must meet a high epistemic standard. If it is significantly probable that one's justification for harming another is unsound, then, prima facie, that behavior is seriously wrong. But if free-will skeptics are right, neither libertarians nor compatibilists satisfy this epistemic standard, and hence individuals do not justly deserve to be punished. And if individuals do not justly deserve to be punished, there is no violation of the rights of victims to deny them the revenge they seek. Even retributivists would acknowledge that the desire for revenge and retribution has its limits. The principle of proportionality, despite its weaknesses, dictates that punishments that are disproportionate to the wrong done (whatever that ultimately amounts to) would be unjustified. Hence, if the victim of an armed robbery wanted to see their perpetrator executed, and this was deemed disproportionate punishment by the standards of retributivism, it would not be a violation of the victim's rights on that theory to prohibit said execution. By extension, if free-will skeptics are right, and retributive punishment itself is unjustified, then to deny victims their desire for revenge would likewise not be a violation of their rights. For victims to have the right to see suffering and harm imposed on their perpetrators, it would have to be the case that such harm was justified. According to free-will skeptics, however, neither victims of violent crime nor the state acting on their behalf is justified in causing more harm than is minimally required for adequate protection.

So, while I think the public health-quarantine model *can* successfully protect a victim's right, we have to be careful not to go beyond what is justified. In the case of Brock and Aaliyah, I maintain that the public health-quarantine model can justify placing liberty-limiting restrictions on Brock, including those that protect Aaliyah from further victimization, but we also have to be careful not to go beyond what is necessary and avoid giving into vengeful desires that seek punishments or sanctions that are themselves unjust. It's a difficult balance, but I think the public health-quarantine model gets it right.

2. Replies to Raff Donelson

I now turn to the critical comments of Raff Donelson. In contrast with those of McCormick, who focuses mainly on the public health-quarantine model, Donelson, though generally agreeing with my ultimate conclusions, focuses mainly on my arguments against retributivism and other nonretributive alternatives. I begin by addressing his concerns about the epistemic argument, pragmatic

worries about free will, and the disjunct of my misalignment and poor epistemic position arguments. I then conclude by addressing Donelson's concerns about desert and the moral education view.

2.1 Epistemic Argument

With regard to the epistemic argument, Donelson writes: 'Like anyone else, I tend to agree that we should refrain from marshalling the resources of the state to punish unless we have strong grounds for issuing the punishment' (XX). Nevertheless, he continues, 'I [still] wonder if there is not a slightly different epistemic argument that one could run which would imperil the point made by Caruso'. While he does not fully spell out this alternative epistemic argument, his concern seems to be that, *if* (1) it is morally obligatory to punish wrongdoers (as Kant suggests), then (2) it would be wrong *not* to punish someone who is really blameworthy. Hence, (3) even in the face of uncertainty about the existence of free will, perhaps the wrongfulness of not giving wrongdoers their just deserts outweigh concerns about unjustly imposing retributive punishment.

There are, however, a number of things to say in response to this concern. First, not everyone agrees with (1). In fact, many (perhaps most) contemporary retributivists themselves reject it—with the exception of, say, Michael S. Moore (1993, 1997). Both *weak* and *moderate retributivists*, for instance, reject the moral obligation to punish wrongdoers. Moderate retributivism, for example, maintains that negative desert is necessary and sufficient for punishment but that desert does not mandate punishment or provide an obligation to punish in all circumstances—that is, there may be other goods that outweigh punishing the deserving or giving them their just deserts. Retributivist Leo Zaibert defends a kind of moderate retributivism when he argues:

> There are many reasons why sometimes refraining from punishing a deserving wrongdoer is more valuable than punishing him—even if one believes that there is [intrinsic] value in inflicting deserved punishment. Perhaps the most conspicuous cases are those in which the refraining is related to resource-allocation and opportunity costs. . . . To acknowledge the existence of these cases is not to thereby deny the value of deserved punishment: it is simply to recognize that this value, like any value, can be—and often is—lesser than other values. (2018: 20)

On this conception of retributivism, which is dominant among contemporary theorists, negative desert grounds a justified reason to punish but not a duty (see, also, Berman 2016). For such moderate retributivists, negative desert is

sufficient to justify punishment, but other values and considerations may outweigh inflicting the deserved punishment.

Second, I challenge Donelson's weighing of risk and unjust wrongdoing in the face of epistemic uncertainty about the existence of free will. Premise 1 of my epistemic argument places the burden of proof on those who want to justify legal punishment, since the harms caused in this case are often quite severe, including the loss of liberty, deprivation, and, in some cases, even death. Victor Tadros spells out these harms:

> Punishment is probably the most awful thing modern democratic states systematically do to their own citizens. Every modern democratic state imprisons thousands of offenders every year, depriving them of their liberty, causing them a great deal of psychological and sometimes physical harm. Relationships are destroyed, jobs are lost, the risk of the offender being harmed by other offenders is increased, and all at great expense to the state. (2011: 1)

Given the gravity of these harms, the justification for legal punishment must meet a high epistemic standard. If it is significantly probable that one's justification for harming another is unsound, then, prima facie, that behavior is seriously wrong (Pereboom 2001: 199; see also Vilhauer 2009).

Support for premise 1 can be found both in the law and everyday practice. As Michael Corrado writes:

> The notion of a burden of proof comes to us from the adversarial courtroom, where it guides the presentation of evidence. In both criminal and civil cases the defendant is presumed not guilty or not liable, and it is up to the accuser to persuade the finder of fact. The only difference between the two cases lies in the measure of the burden that must be carried, which depends upon the seriousness of the outcome. When all that is at issue is the allocation of a loss that can be measured in financial terms, the accuser needs only to prove the defendant's fault by a preponderance of the evidence, but where the defendant's very life or freedom is at stake the burden is considerably higher: the prosecutor must prove beyond a reasonable doubt. (2017: 1)

In the case of legal punishment where the severity of harm is beyond question, I maintain that we should place the highest burden possible upon the state.

If the state is going to punish someone for first-degree murder, say, then the epistemic bar that needs to be reached is guilt beyond a reasonable doubt. But does this burden of proof carry over to theoretical debates (for example,

the debate over free will and moral responsibility)? Here I follow Pigliucci and Maarten (2014) as well as Corrado (2017: 3) in distinguishing between *evidential* burden of proof, which comes into play only when there are no costs associated with a wrong answer, and *prudential* burden of proof, which comes into play precisely when there are significant costs associated with a wrong answer. Corrado applies the distinction to theoretical matters:

> In a purely philosophical contest where nothing of a practical nature hangs on the outcome it is the evidential burden of proof that is required, and the standard of proof must be 'by a preponderance of the evidence': whoever simply has the better evidence must win. On the other hand, if something practical does depend on the outcome of the philosophical debate, then what would matter is the prudential burden. The costs on either side would determine the allocation of the burden and the standard by which satisfaction of the burden is to be measured. (2017: 3)

I contend that given the practical importance of moral responsibility to legal punishment, and given the gravity of harm caused by legal punishment (to the individuals punished as well as those family and friends who depend upon the imprisoned for income, love, support, and/or parenting), the proper epistemic standard to adopt is the prudential burden of proof beyond a reasonable doubt.

Benjamin Vilhauer, for instance, has persuasively argued that 'if it can be reasonably doubted that someone had free will with respect to some action, then it is a requirement of justice to refrain from doing serious retributive harm to him in response to that action' (2009: 131). Derk Pereboom has also proposed applying the reasonable-doubt standard (2001: 161). I maintain that the proof-beyond-a-reasonable-doubt standard is the appropriate epistemic standard to apply when we are talking about intentional harm and institutional punishment. When the stakes are high, as they are with legal punishment, both the law and everyday practice demand that we set the epistemic bar accordingly. As Vilhauer notes, the prudential burden of proof beyond a reasonable doubt has a close kinship to another 'reasonable doubt' principle, which is widely recognized to be a requirement of justice: 'That is the requirement in Anglo-American criminal legal proceedings that the accused can only be convicted of a crime if it is proven beyond reasonable doubt that he acted criminally' (2009: 133). The grounds for accepting this high epistemic standard for criminal conviction are the same as the grounds for accepting it with regard to premise 1 of my Epistemic Argument.

Donelson has given me no reason to think I am wrong about where the burden of proof lies. Nor has he given me reason to think we should shift the burden to those who oppose intentionally harming individuals under conditions of

epistemic uncertainty about the existence of free will. Just as the current criminal law holds that it is seriously wrong to intentionally punish (and hence harm) an individual when there is reasonable doubt about their guilt, I maintain that it is equally wrong to intentionally harm an individual when there is reasonable doubt that they are morally responsible in the basic-desert sense and hence justly deserve such harm.

2.2 *Pragmatic Worry about Free Will*

Donelson's second concern has to do with the pragmatic implications of my free-will skepticism. He writes: 'If we are as Caruso suggests [i.e., byproducts of our histories and circumstances], I find myself unable to see why human beings have any special value that could be the source of duties of respect' (XX). While this is a common concern, it is one I have addressed at great length, both in the book and elsewhere (Caruso 2018, 2019, 2021a, 2021b; Pereboom and Caruso 2018; Dennett and Caruso 2021; see also Pereboom 2001, 2014). In short, I argue that there is no inconsistency in accepting a Kantian regard for respect of persons without accepting Kant's particular attitudes on free will (see Pereboom 2001: 150–52; Vilhauer 2009, 2013a, 2013b; Caruso 2021a). Vilhauer (2009, 2013a, 2013b), for instance, has convincingly argued that there is an important, often overlooked, distinction between 'action-based' desert claims and 'personhood-based' desert claims. The former is the kind of desert claim at issue in the free-will debate while the latter is not. Consider the backward-looking justification of blame and praise, punishment and reward:

> The only sort of reference event that mainstream ethicists typically accept in backward-looking justifications is action—more specifically, actions that they take agents to be morally responsible for performing (if they suppose that there are such things). According to backward-looking justifications of this form, we deserve to be treated in particular ways because of how we have acted. I . . . refer to such justifications as action-based desert claims. Since we can only deserve to be treated in particular ways based on our actions if we are morally responsible, action-based desert claims imply that the agent at issue acted with free will. (Vilhauer 2013a: 149)

Hence, action-based desert claims presuppose the existence of free will.

Personhood-desert claims, on the other hand, arguably do not. This is because we can ground respect for persons in the Kantian principle that *persons must always be treated as ends, and never as mere means*, without adopting Kant's own

particular views on free will. That is, instead of grounding respect for persons in some Kantian notion of 'transcendental freedom', we can ground it in more naturalistic considerations. As Vilhauer explains, free-will skeptics 'can make such dignity and respect for persons a central moral principle if they respect people as rational agents rather than as free agents, and if they regard agents as autonomous not with respect of the laws of nature, but instead with respect to the undue influence of other agents' (148). This is what Vilhauer means by personhood-desert claims:

> We need only look to the people with whom we are interacting to find a basis for desert-claims that constrain consequentialist justifications. Personhood can provide a basis for desert-claims which is irreducibly different from action, and which does not depend upon free will in the way action-based desert does. Persons deserve to be treated only as they would rationally consent to be treated, just because they are persons. (151)

Personhood-based desert claims are therefore distinct from action-based desert claims. And free-will skeptics can recognize and appeal to the former while rejecting the latter. Adopting the skeptical perspective is therefore consistent with a duty to respect persons.

As for Donelson's concerns about autonomy and self-determination, I would reiterate a point I have previously made in response to critics, - that as a free-will skeptic I agree that there are important differences between agents who have the kind of control compatibilists have identified. Such distinctions are in fact undeniable. A normal adult who is responsive to reasons, for instance, differs in significant ways from one who is suffering from brain damage, Alzheimer's, or severe mental illness. And an agent whose effective desire conforms appropriately to their second-order desire for which effective desire they will have differs from an agent who lacks such integration. I have no issue, then, with acknowledging various degrees of control or autonomy—in fact, I think compatibilists have done a great job highlighting these differences. My disagreement has more to do with the conditions required for what I call *basic-desert* moral responsibility. As a free-will skeptic, I maintain that the kind of control and reasons-responsiveness compatibilists point to, though important, is not enough to ground basic-desert moral responsibility, such as the kind of responsibility that would make us truly deserving of blame and praise, punishment and reward in a purely backward-looking sense.

My response, then, to Donelson's concern about the helmsman who is made to steer in a certain direction by a pirate at knifepoint should be obvious. Such an individual would lack the kind of control, autonomy, and self-determination

that a normal, noncoerced, reasons-responsive agent would have. Acknowledging this, however, is no contradiction, nor is it in conflict with my free-will skepticism. One can, as I do, recognize the different levels of control an agent can possess, as well as the causal efficacy of our choices and intentions, *and yet still deny* that such control is equivalent to free will and sufficient (perhaps along with some additional conditions) for basic-desert moral responsibility. Hence, self-determination is consistent with my skepticism since agents can still exercise varying degrees of it *even if* they are not morally responsible in the basic-desert sense for who they are and what they do. That is, even if such capacities give agents a certain kind of control over their decisions, it remains the case that the particular reasons that move us, along with the psychological predispositions, likes and dislikes, and other constitutive factors that make us who we are, ultimately result from factors beyond our control (e.g., determinism, indeterminism, or luck), and I contend that these factors undermine basic-desert moral responsibility.

2.3 On the Disjunct Argument

Donelson's third objection is aimed at the disjunct of my Misalignment Argument and the Poor Epistemic Position Argument. As Donelson summarizes:

> For Caruso, contemporary criminal justice systems currently fail to consider important (mitigating) factors in assigning criminal blame. In doing so, these systems are unjustified; however, if one were to try to modify the systems to take account of those factors, one will find that the state lacks the epistemic resources to apprehend the relevant factors. (XX)

Donelson is not convinced by this argument for the following reason:

> Caruso rightly notes that sentencing courts should be considering things like poverty, unjust deprivation of opportunity, and serious mental illness. Of course, if one were to read sentencing briefs one would see that sentencing judges and juries hear about those factors all the time. If a court hears the mitigation and still issues a sentence that scholars or defense counsel find to be too long, that simply indicates a philosophical disagreement about the mitigating force of particular information. It does not indicate an *inability* of the criminal justice system to consider such factors. (XX)

While I understand Donelson's point, I think it misses the mark. The point of my Misalignment Argument is that retributivism demands that we distribute

punishment in accordance with desert, but it is not at all clear that the criminal law is capable of properly tracking the desert and blameworthiness of individuals in any reliable way. This is because criminal law and judicial procedure is not properly designed to account for all the various factors that affect blameworthiness, and as a result the moral criteria of blameworthiness is often misaligned with the legal criteria of guilt. Sure, judges and juries may hear about various mitigating factors and then dismiss them, but I disagree that this 'simply indicates a philosophical disagreement about the mitigating force of particular information'.

For one, as my Poor Epistemic Argument highlights, for the state to be able to justly distribute legal punishment in accordance with desert, it needs to be in the proper epistemic position to know what an agent basically deserves, but since the state is (almost) never in the proper epistemic position to know what an agent basically deserves, it follows that the state is not able to justly distribute legal punishment in accordance with desert. The problem is not just a philosophical disagreement about how to weigh the mitigating force of particular information. Rather, the state is seriously epistemically and practically compromised in a number of ways. These include being unable to properly track and weigh the extent to which prior abuse and disadvantage should mitigate responsibility; being unable to dedicate the time, resources, and effort needed to accurately assess desert and blameworthiness in most cases; cognitive biases that cloud our judgments; and the general difficulty of knowing how addiction, diminished capacities, social injustice, and other relevant factors should affect judgments of desert. The simple fact is that the state, including judges and juries, are limited in their time, ability, and resources to accurately track and assess issues of desert.

Furthermore, as Erin Kelly correctly notes:

> Conditions that excuse moral failings—such as ignorance, provocation, and mental illness—have limited application in law. This demonstrates a lack of alignment between law and morality. Considerations that mitigate moral blame are often irrelevant to legal findings of criminal guilt. For example, poverty and other unjust deprivations of opportunity have no mitigating relevance in the courtroom; nor do serious mental illnesses such as sociopathy and schizophrenia. Some criminal defendants have diminished moral culpability and others should not be seen as morally blameworthy at all, yet such factors have no bearing on determinations of legal guilt. (2018: 3–4)

The point here is that it's not just a 'philosophical disagreement about the mitigating force of particular information', but rather the law *structurally prohibits* certain mitigating factors as legally exculpatory (see Caruso 2021a: 129–37 for

examples), and this creates a fundamental misalignment between the legal criteria of guilt and the moral criteria of blameworthiness. Hence, if judges and juries are not legally allowed to consider poverty, racism, and diminished capacities short of those that satisfy the extremely narrow confines of the insanity defense as exculpatory or mitigating, then the legal criteria of guilt is *fundamentally* going to be misaligned with the moral criteria of blameworthiness.

For the forgoing reasons, I conclude that if retributivists resist broadening the range of considerations by which legal guilt can be mitigated, they end up with an unacceptable misalignment between the legal criteria of guilt and the moral criteria of blameworthiness. On the other hand, if they attempt to properly align the legal criteria of guilt with the moral criteria of blameworthiness so that that former accurately tracks the various factors that mitigate basic-desert moral responsibility, they will end up confronting epistemic and practical limitations that make it virtually impossible for the state to properly distribute punishment in accordance with desert.

2.4 Our Shifting Desert Judgments

Donelson's fourth objection is concern with our shifting desert judgments. It is aimed at my *Indeterminacy in Judgment Argument*, which maintains that how the state goes about judging the gravity of wrong done, on the one hand, and what counts as proportional punishment for that wrong, on the other, is open to subjective and cultural biases and prejudices and, as a result, the principle of proportionality in *actual practice* does not provide the kind of protections against abuse that it promises. Donelson correctly quotes me as saying, 'There simply is no magic ledger to look to that objectively and impartially spells out a rank order of wrongs in one column and the punishment deserved for each in another' (Caruso 2021a: 142). In response, he writes:

> This argument is puzzling. Either we think we can answer moral questions or we cannot. If we cannot, Caruso's own project of exposing moral flaws in the retributivist framework is doomed. If we can, I fail to see why this specific rank-ordering task presents an insurmountable obstacle. That there is little or no consensus now over the proper rank-ordering or that we now disagree with humans from the distant past is of little *theoretical* consequence, unless we are given to a skepticism that would threaten Caruso too. (XX)

There is, however, something odd about this reply since making *objective and impartial judgments about desert* is not the same thing as *answering moral questions*.

One can deny our ability to successfully do the former while retaining faith in our ability to do the latter.

Furthermore, 'exposing moral flaws in the retributivist framework' is not the same as (1) objectively and impartially rank-ordering wrongs and (2) determining what counts as proportional punishment for each one. Not only do I have less faith than Donelson in our ability to accurately carry out these dual tasks, I also claim that there are fundamental problems confronting judgements of desert. As Julian Lamont has put it, desert is a highly indeterminate concept that 'requires external values and goals to make it determinate' (1994: 45). That is:

> When people make desert-claims they are not simply telling us what desert itself requires. They unwittingly introduce external values, and make their desert-judgments in light of these values. The reason why so many writers have been able to affirm so confidently such a diverse and conflicting set of desert-claims in debates over distributive [and criminal] justice is not because the true conceptual and moral core of desert is so complex and difficult to discern. It is because the true conceptual and moral core of desert allows the introduction of external values and goals. It is the diversity and conflicting nature of these values which explains the diversity and conflicting nature of desert-claims. This is why differences of opinion over what should constitute the desert-base are not going to be solved by examination of desert itself. The differences do not lie at that level, but rather at the level of values. (49)

And radical changes in values and desert judgments need not take hundreds of years to occur. As Alice Ristroph observes, 'Perceptions of deserved penalty (not simply the optimal deterrent or necessary incapacitation, but the deserved penalty) for smuggling dangerous items onto airplanes or violating airport security regulations probably changed dramatically after September 11, 2001' (2006: 1309).

Of course, Donelson is correct that this is more a practical problem than a theoretical one for the retributivist, but I'm not convinced it's wholly untheoretical. Retributivism 'tells us to punish those who deserve it but fails to give any indication of who deserves it and how much they deserve' (Gruber 2010). As Ristroph (2006, 2009) and Aya Gruber (2010) have noted, while there have been attempts to resolve this problem, such as the currently popular 'empirical retributivism', which defines retributive justice with reference to shared social intuitions of what is deserved and how bad certain crimes and criminals are, serious problems still remain. For instance, 'the social intuitionism school is particularly disturbing in light of studies that reveal social intuitions of justice to be largely racialized' (Gruber 2010). In this sense, 'judgments of "desert" may serve as an

opportunity for racial bias to enter the criminal justice system' (Ristroph 2009: 749). Furthermore, in the eyes of ordinary citizens, many criminals fail to get as much punishment as they deserve (Ristroph 2006). Factoring these intuitions into our judgments of what wrongdoers deserve will only lead to more punitive responses.

I'll leave it to retributivists, then, to work out these difficulties. Perhaps Donelson is correct that a more objective and impartial notion of desert can be articulated, but I'm not optimistic.

2.5 *The Moral Education View*

Lastly, Donelson examines my criticisms of moral education theories of punishment and attempts to respond to them. Moral education theories maintain that 'punishment should not be justified as a deserved evil, but rather as an attempt, by someone who cares, to improve a wayward person' (Hampton 1984: 237). Perhaps the best known of these theories is the one proposed by Jean Hampton in 'The Moral Education Theory of Punishment' (1984), although she later abandoned the view in favor of an expressive theory of retribution. According to the theory, punishment is justified if and only if it gets wrongdoers 'to reflect on the moral reasons for that barrier's existence [i.e., the law's prohibition] so that he will make the decision to reject the prohibited action for moral reasons, rather than for the self-interested reason of avoiding pain' (212). According to Hampton, the moral education theory provides a 'full and complete justification' of punishment (209). The theory seeks to reveal that certain actions are unacceptable not only because they are forbidden by law but because they are morally wrong. The goal is not merely to set boundaries of behavior or action but to show that such behavior or actions are not acceptable because they are immoral. As Hampton puts it, 'Wrong occasions punishment not because pain deserves pain, but because evil deserves correction' (209).

Against such views, I argue (following Pereboom) that (1) there are plenty of criminals for whom education would be useless and (2) punishment may do more harm than good (see Pereboom 2001, 2014, 2021).

In response to (1), Donelson notes that 'punishment is not supposed to be a panacea to every instance of wrongdoing' (XX). He proceeds to suggest that 'the moral education proponent can insist on finding nonpunitive ways of dealing with problems that punishment is not designed to fix', That is, 'maybe incapacitation and medical treatment are apt responses in some cases; maybe some other nonpunitive response is in order' (XX). While I welcome this response, note that it strays quite a bit from Hampton's initial claim that the theory provides a 'full and complete justification' (1984: 209) of punishment in cases of

criminal wrongdoing and instead turns it into a pluralistic approach. In fact, Hampton herself later came to reject her moral education theory because, in her own words, there are 'too many criminals on whom such a [moral educative] message would be completely lost; for example, amoral risk-takers, revolutionary zealots, sociopathic personalities' (1992a: 21). As Richard Dagger explains, 'If moral education provides the 'full and complete justification' for punishment that [Hampton] once sought, then it would see that there is no good reason to punish such people, no matter how heinous the crimes they committed' (2011: 7). The core problem is that by justifying punishment on the grounds of moral education and claiming it is the 'full and complete justification', we are left unable to successfully address violent criminals who, for various reasons, may be unable to benefit from such 'moral education'.

Of course, one could, as Donelson suggests, argue that we should retain the punitive component of the theory for those susceptible to improvement by punishment, and, for all the rest, employ something like the public health-quarantine model. But against such a proposal, I would raise four concerns. First, it's unclear to me that the state would be justified in punishing and intentionally harming some wrongdoers while only incapacitating others *for the same offense*, since the former happen to be sensitive to such punishment. Second, it's highly questionable that it's within the proper function of the state to try to morally educate its citizens by intentionally imposing harsh treatment (i.e., state punishment) on them. There has been much criticism of so-called *moral perfectionist* views of the state, and the problem is even greater when the means by which the state attempts to morally improve its citizens comes in the form of punishment. Third, punishment tends to do more harm than good (the second of my initial objection). Hence, it's far from clear that state punishment is the best way to bring about change and moral improvement in wrongdoers (for more on this point, see Pereboom 2021). Lastly, if, as I believe, there are nonpunitive alternatives that are capable of achieving the same end, then I think we should adopt those alternatives. Of course, as Donelson notes, it remains an open empirical question whether I am right, both about the harm punishment produces and the ability to achieve moral development through nonpunitive means. But I'm willing to test that hypothesis.

One final point: Donelson is partially correct when he points out a similarity between the moral education approach and my own public health-quarantine model, with its focus on rehabilitation and reintegration. Both reject retributivism and its focus on desert. Both also seek the moral improvement of wrongdoers. The big difference, however, is that the public health-quarantine model rejects legal punishment altogether, while the moral education approach attempts to justify it.

Legal punishment, we can say, is the intentional imposition of a penalty for conduct that is represented, either truly or falsely, as a violation of a law of the state, where the imposition of that penalty is sanctioned by the state's authority. More precisely, we can say that 'legal punishment consists in one person's deliberately harming another on behalf of the state in a way that is intended to constitute a fitting response to some offense and to give expression to the state's disapproval of that offense' (Zimmerman 2011; see also Boonin 2008; Walen 2014). This is Michael Zimmerman's definition and it is the one I will adopt in my book since I think it best captures the key features of legal punishment. According to this definition, Person A legally punishes Person B if and only if A acts on behalf of the state in such a way that (1) he harms the punishee—this could include imposing an unpleasant penalty or deprivation; (2) this harm is intended by the state; (3) this harm is believed by the state to be fitting—in particular, fitting to the fact, perhaps in conjunction with some other facts, that the punishee is associated with some legal offense; (4) he thereby expresses the state's disapproval both of the offense and of the offender; and (5) he thereby acts in some legal official capacity (20).

I maintain that the public health-quarantine model does not involve punishment in this way and is therefore nonpunitive. When we quarantine an individual with a communicable disease in order to protect people, we are not intentionally seeking to harm or impose a penalty on them. Nor are we communicating disapproval of them as an agent. The same is true when we incapacitate the dangerous criminal in order to protect society. The right of self-defense and prevention of harm to others justifies the limiting or restricting of liberty, but it does not constitute punishment as standardly understood. This is important for several reasons. First, the model demands that we view individuals holistically and that we adopt a preventive approach, one that understands that individuals are embedded in social systems, that criminal behavior is often the result of social determinants, and that prevention is always preferable to incapacitation. Second, after a criminal offense has occurred, courts would need to work with mental health experts, drug treatment professionals, and social service agencies to seek alternatives to incarceration. Third, for those who must be incapacitated, they would need to be housed in nonpunitive environments designed with the purpose of rehabilitation and reintegration in mind. Since most prisons in the United States, United Kingdom, and Australia are inhospitable and unpleasant places designed for punitive purposes, we would need to redesign them so that the physical environments and spaces we incapacitate people in better serve the goal of rehabilitation and reintegration (see Caruso 2021a: chs 7 and 8). Lastly, with regard to sentencing, voter disenfranchisement, and three-strikes laws, the public health-quarantine model would require radical changes to our current, often excessively punitive, system.

If Donelson wishes to resurrect the moral education justification of punishment, he must overcome the four serious concerns mentioned above as well as explain why such an approach would be preferrable to my nonpunitive alternative. Without a concrete proposal to judge, I will continue to maintain that the public health-quarantine model remains the best nonretributive alternative.

References

Berman, Mitchell. 2016. 'Modest Retributivism', in *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore*, ed. by Kimberly Kessler Ferzan and Stephen J. Morse, (New York: Oxford University Press), pp. 35–48

Boonin, D. 2008. *The Problem of Punishment* (New York: Cambridge University Press)

Caruso, Gregg D. 2012. *Free Will and Consciousness: A Determinist Account of the Illusion of the Free Will* (Lanham, MD: Lexington Books)

Caruso, Gregg D. 2016. 'Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model', *Southwest Philosophical Review*, 32: 25–48

Caruso, Gregg D. 2020. 'Justice with Retribution: An Epistemic Argument Against Retributive Criminal Punishment', *Neuroethics*, 13: 13–28

Caruso, Gregg D. 2021a. *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (New York: Cambridge University Press)

Caruso, Gregg D. 2021b. 'Retributivism, Free Will Skepticism, and the Public Health-Quarantine Model: Replies to Kennedy, Walen, Corrado, Sifferd, Pereboom, and Shaw', *Journal of Legal Philosophy*, 46: 161–215

Caruso, Gregg D. 2021c. 'Rejecting Retributivism: Reply to Leo Zaibert', *Philosopher*, 109: 118–26

Caruso, Gregg D., and Derk Pereboom. 2020. 'A Non-Punitive Alternative to Punishment', in *Routledge Handbook of the Philosophy and Science of Punishment*, ed. by Farah Focquaert, Bruce Waller, and Elizabeth Shaw (New York: Routledge), pp. 355–65

Caruso, Gregg D., and Derk Perenoom. Forthcoming. *Moral Responsibility Skepticism* (New York: Cambridge University Press)

Corrado, Michael L. 2017. 'Punishment and the Burden of Proof', *UNC Legal Studies Research Paper* <https://ssrn.com/abstract=2997654>

Dagger, R. 2011. 'Jean Hampton's Theory of Punishment: A Critical Appreciation', *APA Newsletter on Philosophy of Law*, 10: 6–11

Dennett, Daniel C., and Gregg D. Caruso. 2021. *Just Deserts: Debating Free Will* (New York: Polity Press)

Donelson, Raff. 2022. 'Comments on Gregg Caruso's *Rejecting Retributivism*', *Journal of Practical Ethics* (this issue)

Fischer, John M. 2007. 'Compatibilism', and 'Response to Kane, Pereboom, and Vargas', in *Four Views on Free Will*, ed. by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas (Oxford: Blackwell Press), pp. 44–84, 184–90

Fischer, J. M., and M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press)

Floud, Jean E., and Warren Young. 1981. *Dangerousness and Criminal Justice* (Portsmouth, NH: Heinemann)

Foubert, John D., Angela Clark-Taylor, and Andrew F. Wall. 2020. 'Is Campus Rape Primarily a Serial or One-Time Problem? Evidence from a Multicampus Study', *Violence Against Women*, 26: 296–311

Gruber, A. 2010. 'The False Promise of Retributive Proportionality', *Jotwell Journal*, 12 July <http://scholar.law.colorado.edu/cgi/viewcontent.cgi?article=1220&context=articles>

Hampton, Jean. 1984. 'The Moral Education Theory of Punishment', *Philosophy and Public Affairs*, 13: 208–38

Kelly, Erin. 2018. *The Limits of Blame* (Cambridge, MA: Harvard University Press)

Lamont, J. 1994. 'The Concept of Desert in Distributive Justice', *Philosophical Quarterly*, 44: 45–64

Lisak, David, and Paul Miller. 2002. 'Repeat Rape and Multiple Offending among Undetected Rapists', *Violence and Victims*, 17: 73–84

Lemos, John. 2018. *A Pragmatic Approach to Libertarian Free Will* (New York: Routledge)

McCormick, Kelly. 2022. 'Comments of Gregg Caruso's *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*', *Journal of Practical Ethics* (this issue).

McKenna, Michael. 2012. *Conversation and Responsibility* (New York: Oxford University Press)

McKenna, Michael. 2020. 'Punishment and the Value of Deserved Suffering', *Public Affairs Quarterly*, 34: 97–123

Moore, M. S. 1993. *Act and Crime: The Philosophy of Action and Its Implications for Criminal Law* (New York: Oxford University Press)

Moore, M. S. 1997. *Placing Blame* (New York: Oxford University Press)

Morris, S. 2018. 'The Implications of Rejecting Free Will: An Empirical Analysis', *Philosophical Psychology*, 31: 299–321

Morse, Stephan J. 1999. 'Neither Desert nor Disease', *Legal Theory*, 5: 265

Nadelhoffer, Thomas, S. Bibas, S. Grafton, K. A. Kiehl, A. Mansfield, W. Sinnott-Armstrong, and M. Gazzaniga. 2012. 'Neuroprediction, Violence, and Law: Setting the Stage', *Neuroethics*, 5: 67

Pereboom, Derk. 2001. *Living Without Free Will* (New York: Cambridge University Press)

Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life* (New York: Oxford University Press)

Pereboom, Derk. 2021. *Wrongdoing and Moral Emotions* (New York: 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*, ed. by Gregg D. Caruso and Owen Flanagan (New York: Oxford University Press), pp. 193–222

Pigliucci, Massimo, and Maartin Boudry. 2014. 'Prove It! The Burden of Proof Game in Science v. Pseudoscience Disputes', *Philosophia*, 42: 487–502

Ristroph, A. 2006. 'Desert, Democracy, and Sentencing Reform', *Journal of Criminal Law and Criminology*, 96: 1293–1352

Ristroph, A. 2009. 'How (Not) to Think like a Punisher', *Florida Law Review*, 61: 727–49

Robinson, Paul H. 2008. *Distributive Principles of Criminal Law: Who Should Be Punished How Much* (New York: Oxford University Press)

Sartorio, Carolina. 2016. *Causation and Free Will* (Oxford: Oxford University Press)

Tadros, Victor. 2011. *The Ends of Harm: The Moral Foundations of Criminal Law* (New York: Oxford University Press)

Vilhauer, Benjamin. 2009. 'Free Will and Reasonable Doubt', *American Philosophical Quarterly*, 46: 131–40

Vilhauer, B. 2013a. 'The People Problem', in *Exploring the Illusion of Free Will and Moral Responsibility*, ed. by G. D. Caruso (Lanham, MD: Lexington Books) pp. 141–160

Vilhauer, B. 2013b. 'Persons, Punishment, and Free Will Skepticism', *Philosophical Studies*, 162: 143–63

Walen, Alec. 2014. 'Retributive Justice', *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/entries/justice-retributive/> [accessed 24 November 2022]

Zaibert, Leo. 2018. *Rethinking Punishment* (New York: Cambridge University Press)

Zimmerman, Michael J. 2011. *The Immorality of Punishment* (Peterborough: Broadview Press)