Journal of Legal Philosophy, Vol. 46 No. 2, 2021, pp. 161–215

Retributivism, free will skepticism and the public health-quarantine model: replies to Corrado, Kennedy, Sifferd, Walen, Pereboom and Shaw

Gregg D Caruso
SUNY Corning and New College of the Humanities (NCH London)

I would like to begin by thanking Michael Corrado, Chloë Kennedy, Katrina Sifferd, Alec Walen, Derk Pereboom and Elizabeth Shaw for their astute and challenging comments on my book, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*.[1] It is seldom that one gets the opportunity to put their views to the test by responding to six of the leading figures in their field. While I have had only the briefest time to consider their comments, and more prolonged reflection would no doubt yield more insights, I have already benefited greatly by wrestling with their perceptive criticisms. In this article, I outline the objections, suggestions and critical points presented by each commentor and respond to each as best I can. While I dedicate more space to some challenges than others, this is not a reflection of the quality of the commentaries but is instead due to a limitation on time and space. There is also occasional overlap between the commentaries, and it makes more sense to address common criticisms only once. I begin by responding to Michael Corrado and then proceed in the order indicated in the subtitle.

1 MICHAEL LOUIS CORRADO

Corrado has been a long-standing critic of the public health-quarantine model, despite the fact we agree on several of the core philosophical issues. We both, for instance, consider ourselves free will skeptics. We also agree that the retributive justification of legal punishment must be rejected. One would think, then, that we would end up in a similar place, but that's not the case. Whereas I opt for a nonpunitive approach I call the *public health-quarantine model*, Corrado opts for an account he calls *correction* – which involves 'intentional harsh treatment' aimed at changing behaviour.[2] According to Corrado, intentional harsh treatment is justified, not on the grounds of desert, but by the 'benefit it offers the one corrected'.[3] While correction is consistent with

1. GD Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (Cambridge University Press 2021).
2. ML Corrado, 'Why Do We Resist Hard Incompatibilism? Thoughts on Freedom and Punishment', in Thomas Nadelhoffer (ed), *The Future of Punishment* (Oxford University Press 2013); ML Corrado, 'Fichte and the Psychopath: Criminal Justice Turned Upside Down', in E Shaw, D Pereboom and GD Caruso (eds), *Free Will Skepticism in Law and Society* (Cambridge University Press 2019).
3. Corrado, 'Why Do We Resist Hard Incompatibilism?' (n 2).

Journal compilation © 2021 Edward Elgar Publishing Ltd
The Lypiatts, 15 Lansdown Road, Cheltenham, Glos GL50 2JA, UK
and The William Pratt House, 9 Dewey Court, Northampton MA 01060-3815, USA

free will skepticism, Corrado also believes it retains the protective features of punishment. In a series of recent papers, Corrado has presented several challenging objections to my account[4] – all of which I have taken seriously and have responded to at length.[5] In fact, I dedicate several sections of *Rejecting Retributivism* to responding to Corrado's numerous concerns.

In his commentary here, Corrado begins by arguing that the failure of retribution should not be mistaken for a failure of the Classical Model. The Classical Model, Corrado explains, was developed during the Enlightenment through Cesare Beccaria and maintains that: The state may use force to 'control human behaviour' but 'only against individuals who have violated laws set down in advance by the people's representatives, and only to an extent proportional to the harmfulness of the behaviour, an amount also to be decided in advance by the legislature'.[6] According to Corrado, the Classical Model must be distinguished from the retributive justification that philosophers 'attempted to fit it up with'. With this distinction in place, Corrado defends the following thesis:

> My position in this is (with Caruso) that retribution doesn't succeed as a justification for the classical approach; that it may be that the classical approach, insofar as it depends upon interfering in the lives of individual offenders, can never be fully justified; but (against Caruso) that it is nevertheless morally superior to any of the alternatives that have been proposed, including Quarantine.[7]

While Corrado agrees with me that the current system is 'deplorable', he also maintains that 'it would be a mistake of the first order to suppose that the existing system of crime control is an example of the Classical Model'. Instead of rejecting the Classical ideal in favour of the public health-quarantine model, Corrado argues that it is imperative that we preserve the (putative) protections of the Classical Model – especially its commitment to proportional punishment.

With this as his motivation, Corrado goes on to offer three main objections to the public health-quarantine model, which he refers to simply as the 'Quarantine Model'. First, he objects that, 'the edges of [the] model simply cannot be softened in the way [Caruso] would like to soften them. Both pre-emptive detention and indefinite detention are essential and ineliminable features of it'. Second, he argues that 'the justification for the model, self-defence and the defence of others, doesn't seem workable and in particular doesn't seem to limit the state's use of detention in the ways he suggests'. Lastly, Corrado objects that 'since we do not get to pick the political setting in which our proposals land, a system of broad discretion such as Professor Caruso proposes is dangerous. It leaves the door wide open to the very sorts of excess that motivated Beccaria to set down his model in the first place'.

4. ML Corrado, 'Punishment and the Burden of Proof' (2017) UNC Legal Studies Research Paper (Available at: <https://ssrn.com/abstract=2997654)>; ML Corrado, 'Criminal Quarantine and the Burden of Proof' (2019) 47 Philosophia 1095; Corrado, 'Fichte and the Psychopath' (n 2).
5. See, e.g., D Pereboom and GD Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life', in O Flanagan and GD Caruso (eds), *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience* (Oxford University Press 2018).
6. ML Corrado, 'The Limits of the State's Power to Control Crime' (2021) The Journal of Legal Philosophy (this issue).
7. ibid.

I will argue below that my model can successfully deal with each of these objections. I begin with Corrado's concerns about pre-emptive detention and indefinite detention.[8]

1.1 Indefinite Detention and Pre-emptive Detention

Michael Corrado has argued that one advantage retributive punishment (and his own theory of correction) has over the public health-quarantine model is that the former protects 'the right to be held by the state only for a limited period of time',[9] whereas the latter allows for the possibility of indefinite detention if an individual is determined to be a continued serious threat to society. According to Corrado, the methods of quarantine allow for detention and other restrictions of freedom, 'but impose no limits upon those restrictions except those that further the protection of the community'.[10] He goes on to argue that 'quarantine means a complete and unlimited surrender of autonomy to the state' since the person subject to quarantine or incapacitation 'has no protection against the power of the state' – that is, '[i]f he should remain dangerous after treatment he may be detained indefinitely'.[11]

In response, I would argue, first, that while I do acknowledge that on my model seriously violent criminals who cannot be rehabilitated and remain serious threats would be indefinitely detained, this will not amount to a 'complete and unlimited surrender of autonomy to the state', nor would it result in more people serving out their lives in prison. Second, I argue that the putative advantage Corrado claims retributive punishment (and correction) has is more apparent than real. This is because the principle of proportionality is inherently vague and indeterminate, and as a result it is consistent with the fact that one of every seven people currently in prison in the United States is serving life or virtual life sentences even though the American criminal justice system has long been committed to limiting retributivism and determinate sentencing. Third, I maintain that the rare cases where the public health-quarantine model would favour continuing to hold a violent criminal who cannot be rehabilitated are ones its critics, and the current law, would largely agree on. Furthermore, the conditions of detention for such individuals would be very different on my model from those vile conditions of the US prison system. Lastly, I argue that there are policy solutions available, such as setting the maximum prison term at 20 years but allowing for incremental extensions if absolutely necessary for public safety, as Norway does, as well as placing the burden of proof on the state to evaluate and establish, at regular intervals, that the threat posed by an offender warrants continued incapacitation. Let us take each of these points in turn.

First, there is no reason to think the public health-quarantine model amounts to a 'complete and unlimited surrender of autonomy to the state'. This is because the model places a number of important constraints on the treatment of individuals, including the *principle of least infringement*, the *prohibition on manipulative use*, the *principle of normality*, the *principles of beneficence* and *nonmaleficenc*e, and concern for the well-being of offenders.[12] Even in conditions of incapacitation, my model

8. I should note that many of the replies I offer here are drawn from the book, since Corrado's objections are similar to ones he has already raised and I have already responded to.
9. Corrado, 'Criminal Quarantine and the Burden of Proof' (n 4) 1105.
10. ibid 1106.
11. ibid 1107.
12. See the book for statements and defenses of each of these principles.

requires that the state respect human dignity by preserving the rights of offenders to vote, to not be dehumanized, and to be afforded as much autonomy and liberty as possible consistent with the minimum restrictions needed to protect public safety. According to the principle of normality, for instance, the only right an offender should lose when they are incapacitated is their liberty; they retain all their other rights. These include the right to vote (even while in prison), to health care, to attend school, to phone family and friends, etc. On almost every one of these points, our current methods of punishment sacrifice the autonomy of offenders more completely to the state. By limiting his concerns to indefinite detention, Corrado obfuscates this point and completely overlooks the fact that retributive punishment sacrifices more of an offender's rights to the state than the public health-quarantine model.

Second, with regard to the number of people who will never end up being released, which seems to be Corrado's driving concern with regard to indefinite detention, a strong case can be made that the public health-quarantine model far and away has the advantage here. Currently, one in seven people in prison in the United States is serving life or 'virtual life' sentences, that is, sentences with a term of years that exceed an individual's natural life expectancy. These sentences are determinate in length and, at least conceivably, within the realm of what is permitted by the principle of proportionality. Of course, a retributivist could argue that the number of people serving life and virtual life sentences is excessive. But I argue in the book that the principle of proportionality is vague, elastic and indeterminate when it comes to judging gravity and proportionality, hence it is virtually impossible to determine precisely what the upper limit of punishment should be for a violent crime. Findings compiled by the Justice Policy Institute, for instance, have found that sentencing times vary widely for the same crime across the globe. And the fact that a prison term is of determinate length, or specifies an acceptable range, by no means guarantees offenders will ever be released. In fact, of those serving a life sentence in the United States, about a third are without the chance of parole. Of the remainder, political considerations have made it increasingly difficult to secure parole in many states. In addition, a large number of offenders are serving 'virtual life' sentence. For example, a 40-year prison term imposed on a 35-year-old offender essentially equates to life imprisonment. And it is not uncommon for punitive systems to hand out impossible sentences, like multiple life sentences, to make a point or because they believe them to be proportional to the wrong done. Terry Nichols, for example, one of the accomplices in the 1995 Oklahoma City bombing, was given 161 life sentences with no possibility for parole.

In contrast, I maintain that there are good reasons to think that implementing the public health-quarantine model will actually result in a drastic reduction in the number of people incapacitated and the number of people held indefinitely. This claim is supported by a number of different considerations. A recent study by the Brennan Center for Justice found, for instance, that 39 percent of the US prison population (roughly 576,000 people) are behind bars with little public safety rationale.[13] According to the report, up to 25 percent of prisoners (364,000 people), almost all nonviolent, lower-level offenders, would be better served by alternatives to incarceration such as treatment, community service and probation. They estimate that another 14 percent (212,000 people) have already served long sentences and can be safely set free with little or no risk to public safety. If these numbers are correct, adopting the public

13. J Austin, L-B Eisen, J Cullen and J Frank, 'How Many Americans Are Unnecessarily Incarcerated?' Brennan Center for Justice at New York University School of Law (2016). Available at: <www.brennancenter.org/sites/default/files/publications/Unnecessarily_Incarcerated_0.pdf>.

health-quarantine model would result in more than a half million people being released from prison with little to no increased risk to public safety.

Research also indicates that the risk of recidivism drops dramatically as prisoners get older, and there will come a time when an additional year of prison no longer yields a meaningful reduction in the risk of recidivism among older prisoners.[14] For instance, research by the United States Sentencing Commission has found that older offenders are 'substantially less likely than younger offenders to recidivate following release'.[15] So, while a retributivist could maintain that a 65- or 75-year-old should continue to be incarcerated because they deserve it, the public health-quarantine model would need to consider the chances of recidivism and the kind of harm they pose to society. Since the majority of elderly prisoners pose no serious public safety concerns, the public health-quarantine model would recommend release, with perhaps additional assistance and supervision. Of course, there will be occasional cases where an individual will need to be incapacitated indefinitely, but these will be the exception, not the norm. Given these considerations, I maintain that adopting the public health-quarantine model would drastically reduce, not increase, the number of people serving out their lives in prison.

Third, I maintain that there are additional policy solutions available for the public health-quarantine model to adopt, and these can provide additional protections against potential abuse. The public health-quarantine model already maintains the principle of least infringement, which requires that we always seek alternatives to incapacitation wherever and whenever possible. But beyond this, I also recommend establishing an upper limit of around twenty years in prison as a maximum penalty for the most dangerous of violent criminals, as Norway does, but to also allow for exceptional and incremental extensions when, say, a serial rapist or mass murderer has not been rehabilitated and continues to pose a serious threat to public safety. This upper limit allows for rehabilitated offenders to be released at any point prior to the end of their term, but it prohibits handing down sentences longer than 20 years as a judicial option. I would also recommend placing the burden of proof on the state to establish, at regular intervals, that the threat posed by an offender warrants continued incapacitation. If the state cannot satisfy that burden of proof with documentation – for example, evidence of aggression and violence in prison, credible threats to continue to do harm upon release, etc. – then the offender should be released.

While some might think these recommendations unrealistic, it is important to recognize that sentences of more than 20 years are quite rare in many democratic nations. Additionally, most offenders age out of crime at a certain point, while others mature, educate and rehabilitate themselves – especially when they are given support, counselling and opportunity. I maintain that it is inhumane and contrary to the demands of public safety to continue to incapacitate such individuals. On the other hand, retributivists who favour life or virtual life sentences as deserved proportional punishment end up leaving offenders completely and utterly powerless and without the chance to alter a sentence that offers them no possibility of release or rehabilitation. This amounts to a 'complete and unlimited surrender of autonomy to the state',

14. See K Kim and B Peterson, *Aging Behind Bars: Trends and Implications of Graying Prisoners in the Federal Prison System* (Urban Institute 2014); United States Sentencing Commission, *The Effects of Aging on Recidivism Among Federal Offenders* (2017) United States Sentencing Commission.

15. United States Sentencing Commission, *The Effects of Aging on Recidivism Among Federal Offenders* (n 14) 3.

but in the exact opposite direction than Corrado would have us believe. I therefore contend that the public health-quarantine model actually provides more protections and opportunities for offenders to be released than retributivism.

Finally, the kinds of cases where the public health-quarantine model would allow for indefinite incapacitation or quarantine are intuitively cases most would agree are justified.[16] We also need to recognize that existing punitive approaches, like retributivism, essentially allow for the same possibility in practice. Take, for example, the case of Andres Breivik who killed 77 people in 2011, the worst case of mass murder in Norway's history. He was given the maximum sentence allowed in Norway, which is only 21 years. At the end of that period, the state can add five more years of civil incapacitation on the grounds of public safety if he has not been rehabilitated. It is conceivable that this can be done over and over again resulting in indefinite detention. Corrado might object that this violates 'the right to be held by the state only for a limited period of time'. But what, exactly, would a retributivist (or Corrado) do differently? In a case like Breivik's, there are generally two sentencing options available to retributivists (assuming he is judged competent): They could sentence him to life without the chance of parole right from the start, putting an end to his chances of ever being released, or they could give him a sentence that allows for the chance of parole after serving a minimum number of years in prison.

The first option, I have argued, is inhumane since it precludes from the outset the possibility of rehabilitation, violates the principle of least infringement, discourages the state from working to rehabilitate offenders, and prevents the reassessment of individual cases as circumstances change. The second option, however, provides essentially the same possibility of indefinite detention since the parole board will have the power and discretion to determine whether Breivik should be released or not. If the board decides he should not be released, he will continue to be imprisoned. The key difference, however, between the retributivist approach and my own is that my approach (a) eliminates the option of life without parole, allowing for the possibility of rehabilitation and release, (b) establishes a maximum upper limit that will apply in all non-exceptional cases, (c) places the burden of proof on the state to justify any and all exceptions to that maximum, and (d) requires that no one is incapacitated longer than is absolutely necessary to protect public safety. My model also requires that offenders be housed in humane conditions and treated in accordance with the principle of normality, respect for persons and concern for human well-being.

For the preceding reasons, I maintain that the public health-quarantine model does not amount to a 'complete and unlimited surrender of autonomy to the state', nor will it result in more people serving out their lives in prison. While it does allow for the possibility of indefinite detention in exceptional cases, which is not necessarily a bad thing, most criminal offenders will 'age out' of crime or pose less of a threat over time, allowing for less restrictive measures to be taken to protect public safety. The public health-quarantine model also places several restrictions and humanitarian consideration on the treatment of offenders, which gives it a distinct advantage over current retributive policies.

Moving on then to the issue of pre-emptive detention, Corrado is rightly concerned about the prospect of incapacitating those who pose threats but have not yet committed a crime. In response to this concern, I offer several reasons in the book for opposing the pre-emptive incapacitation of competent, reasons-responsive agents in

16. See, for instance my discussion of 'Typhoid Mary' starting on page 336 in *Rejecting Retributivism.*

practically all real-world cases. 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 as Stephen Morse points out, detaining someone on the basis of a screening method that frequently yields false positives is seriously morally objectionable.[17]

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[18] 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.

These considerations, I contend, will block pre-emptive 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.[19]

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 provides 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

17. SJ Morse, 'Neither Desert Nor Disease' (1999) 5 Legal Theory 265; T Nadelhoffer, S Bibas, S Grafton, KA Kiehl, A Mansfield, W Sinnott-Armstrong and M Gazzaniga, 'Neuroprediction, Violence, and Law: Setting the Stage' (2012) 5 Neuroethics 67.
18. JE Floud and W Young, *Dangerousness and Criminal Justice* (Heinemann 1981).
19. Pereboom and Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life' (n 5) 215.

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 pre-emptive detention.

Corrado, himself, acknowledges that he would also detain this individual for the week. In explaining why, he argues that we could view the drug as undermining the agent's reasons-responsiveness and that in such cases it is sometimes permissible to pre-emptively incapacitate non-reasons-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 other.

I agree with Corrado that 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. And I agree with Corrado that the justification for such commitment lies largely in the fact that such individuals 'are not just dangerous but both dangerous and unable to conform their behaviour to the law (whether from a deficit of reason or of will)'.[20] I see no reason, however, 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 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 non-maleficence, 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 favour of allowing individuals the liberty to commit criminal acts before incapacitation can be justified. Moderately reasons-responsive agents are able to conform their behaviour to the reasons they have for action.[21] 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

20. ML Corrado, 'Criminal Quarantine and the Burden of Proof' (n 4) 1098.
21. See, e.g., JM Fischer and M Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press 1998).

countervailing moral considerations that must be taken into account. I therefore argue that the pre-emptive incapacitation of competent, reasons-responsive agents should be prohibited in practically all real-world cases. 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 pre-emptive measures should be taken.

Lastly, I would also suggest that in many real-world cases concerns about pre-emptive incapacitation are not pertinent. That is because the actions and behaviours 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 pre-emptive. Instead, it would be a reaction to criminal offenses already performed and hence justified on the right of self-defence. 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 behaviour. 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 pre-emptive; 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 pre-emptive 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 behaviour, but from sociological and/or neuroscientific data indicating an increased risk of violence. But in the book I argue that such data are generally insufficient to overcome the presumption of harmlessness and the epistemic burden of proof.

1.2 Self-Defense and Detention

Corrado's second objection is to challenge the very foundation of the quarantine analogy by arguing that, 'we do not have an uncontroversial right to quarantine those with serious communicable diseases, but only those with *certain* serious communicable diseases – those diseases that it is not within the power of the afflicted person to control in response to the threat of punishment'. He provides the following example based on HIV: 'we do not quarantine those with certain very serious, life-threatening, easily communicated sexual diseases when their transmission can be adequately constrained by the threat of punishment'. While I agree with Corrado that an agent's *ability to control* the spread of their communicable disease is relevant to justifying liberty-limiting restrictions, I do not see why this is a serious problem for my account or for the justification of quarantine in general.

The right of self-defence and defence of others, I maintain, allows for the quarantine of those individuals with serious communicable diseases when, and only when, there is a serious threat to public health and safety and no less restrictive measures are available. Of course, not all threats to public health and safety warrant quarantine. Those communicable diseases that do *not* pose a significant threat to public health, *and/or* can be prevented by reasonable measures on the part of the individual, do not warrant quarantine since less restrictive measures are available in those cases. Furthermore, understanding the nature of the communicable disease and how it is transmitted is important in assessing the level of risk it poses to public health and whether quarantine is warranted for those who contract it. For cases like HIV, where it is within the control of competent and reasons-responsive agents to effectively prevent its spread by, say, abstaining from unprotected sex and the use of unclean needles, I would argue that considerations of liberty and the principle of least restriction count decisively against the use of quarantine. Carriers of HIV should be afforded the presumption of harmlessness since they have the ability to prevent the spread of the virus (or, more accurate, reduce the risk of transmission to an acceptably low level) by taking the necessary precautions. And this ability in no way presupposes the kind of free will I deny.[22]

The distinction Corrado draws therefore has nothing to do with retaining the 'threat of punishment', but rather with the different levels of threat each disease poses and whether less restrictive measures are available to effectively control its transmission. In this way, the agent's *ability to control* the spread of their communicable disease is indeed relevant, but not because reasons-responsiveness justifies punishment or is sufficient to ground basic desert. For some reason, Corrado seems stuck on the idea that my model is unable to acknowledge the importance of reasons-responsiveness, but that is simply not true. Neither free will skepticism nor the public health-quarantine model implies that the difference between agents who are reasons-responsive and those who are not is irrelevant. In fact, throughout the book I repeatedly stress the fact that the difference is crucial in at least the following three ways: (a) assessing the risk an individual poses moving forward, (b) determining which preventative measures and interventions will be most effective, and (c) deciding which rehabilitative measures to adopt.

Moving on, then, to Corrado's second concern with my appeal to the right of self-defence and defence of others, he argues that, 'I have no clear idea exactly where the moral boundaries are and I don't think anyone else has, either'. He provides the following example: 'Suppose that great deal of harm could be prevented by threatening a potential offender's innocent family or neighbours … I cannot see how the Quarantine Model, insofar as it rests upon a defensive basis, can avoid the kind of counterexample – the "manipulative use" counterexample – that utilitarianism is subject to'. This objection, however, overlooks several important points I make in the book. For instance, in Chapter 6, I propose the following *conflict resolution principle*, designed to deal with conflicts between public health and safety (on the one hand) and individual liberty and autonomy (on the other). It states:

> **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 (a) in accordance with the right of self-defence and the prevention of harm to others, where (b) this right of self-defence is applied to an individual threat and is calibrated to the danger posed by that threat (not some unrelated threat), and (c) 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.

The first condition (a) maintains that liberty can be limited in cases of quarantine and incapacitation but only in accordance with the general right of self-defence and defence of others. This is significant because it distinguishes the public health-quarantine model from more general utilitarian and consequentialist approaches to punishment. Rather

22. Of course, if the HIV patient *refuses* to take the necessary precautions – just as Typhoid Mary refused to wash her hands or give up working as a cook – then, of course, the right of self-defense and defense of others could justify quarantining them.

than appealing to an increase in some aggregate good (e.g., pleasure) or the benefits of general deterrence as a liberty-limiting justification, the conflict resolution principle states that liberty can only be limited in accordance with the right of self-defence and the defence of others. This, I contend, avoids the kinds of objections typically raised against consequentialist theories – for example, cases where framing innocent individuals or using harsh and severe punishment would be the most effective way to deter crime.

Consider the example of framing an innocent person to prevent a riot.[23] While consequentialists may argue against the practicality of engaging in such practices, claiming that it would erode important protections and produce poorer outcomes in the long run, they must nonetheless allow for the possibility of exceptions. Furthermore, this reply remains insensitive to the fundamental unfairness of punishing an innocent person, pointing only to practicality as a reason to avoid it. While some critics worry that the public health-quarantine model would likewise allow for the framing of an innocent person, the conflict resolution principle reveals why this is not the case.

First and foremost, innocent people do not pose a threat to society, and as a result the right of self-defence would not justify incapacitating them. Second, condition (b) of the conflict resolution principle states that the right of self-defence and prevention of harm to others only apply to the danger posed by individualized threats, not general threats. That is, the right of self-defence only applies to the source of an individual threat and must be calibrated to the danger posed by that threat, not some unrelated threat. Accordingly, it would be wrong to incapacitate an innocent person because that person, being innocent, is not a danger to society nor are they the source of the threat posed by the impending riot. To limit the liberty of an individual because of concerns about the safety of society that emanate from a different source would be a violation of the conflict resolution principle and any intuitive understanding of the right of self-defence. The same is true of 'threatening a potential offender's innocent family or neighbours' in Corrado's example.

1.3 Potential for Political Abuse

Corrado's last and final objection is that, 'crime and the use of state violence to control it are much too serious a matter to be left to the discretion (read 'arbitrariness') of whoever happens to be in charge'. Perhaps it is, but isn't this a general argument against giving the state the power to punish and intentionally harm its own citizens? As Victor Tadros correctly notes:

> 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 offenders being harmed by other offenders is increased, and all at great expense to the state.[24]

The *problem of punishment*[25] is made even worse by the fact that state punishment involves the *intentional* and *deliberate* harming of the punishee in a way that is

23. See HJ McCloskey, 'A Non-Utilitarian Approach to Punishment', in G Ezorsky (ed), *Philosophical Perspectives on Punishment* (State University of New York Press 1972).
24. V Tadros, *The End of Harm: The Moral Foundations of Criminal Law* (Oxford University Press 2011) 1.
25. D Boonin, *The Problem of Punishment* (Cambridge University Press 2008).

intended to constitute a fitting response to some offense and to give expression to the state's disapproval of that offense.[26]

If Corrado is genuinely concerned about political abuse when it comes to the 'use of state violence to control [crime]', wouldn't it be wiser to adopt a *non-punitive* approach, like my own, that does not seek to justify punishment (traditionally understood) or the intentional infliction of suffering/harm/harsh treatment by the state? In fact, retributivism and the Classical Model has proven itself extremely vulnerable to abuse by 'tough on crime' politicians, lawyers and judges, leading to our current 'mass incarceration' crisis. Consider the fact that the US criminal justice system has long embraced the retributive justification of legal punishment, the prohibition on punishing innocent people, and the proportionality requirement, yet this has not prevented the mass incarceration of its citizens nor has it been sufficient to protect against disproportionate punishment. With only 4.5 percent of the world's population, the United States imprisons 25 percent of the world's prisoners – far more than any other nation in the world. And not only does the United States imprison at a much higher rate, it also imprisons in notoriously harsh conditions and its sentences tend to be more punitive than other countries.

While I'm not suggesting that these problems are due solely to retributivism – since there is good reason to think that 'three strikes' laws, mandatory minimums and other policies responsible for mass incarceration were also motivated by the desire for deterrence my point is simply that commitment to limited retributivism and the Classical Model has not *prevented* these abuses. Corrado seems aware of this when he writes: 'Of course, there are certain forms of arbitrariness that no model, not even the Classical Model, can rule out: there is no way to deprive a democratic legislature of the power to determine what sort of behaviour should be criminalized and for how long. The most we can do is to persuade our legislators to avoid abusing their discretion in those matters'. Nevertheless, he seems to think that my model is uniquely susceptible to abuse and arbitrariness.

In response, I would argue three things. First, Corrado is simply mistaken about the putative protections provided by the principle of proportionality. As I argue in Section 4.3 of the book, how the state goes about judging and ranking such things as *gravity of wrong*, on the one hand, and what counts as *proportional punishment*, on the other, are wide open to subjective biases and prejudices. 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 proportional punishments for each in the other. This is obvious from that fact that retributivists often disagree with one another about how to measure each. If the history of punitive practices and institutions has taught us anything, it is that judgments of what counts as a grievous wrong are hypersensitive to cultural biases, prejudices and power relations. And even when there is wide agreement on the gravity of a wrong, there is still often disagreement about what kind of punishment is deserved or proportionate. For instance, all retributivists can agree that intentionally murdering an innocent person is a grievous wrong, but they can, and often do, disagree on what counts as 'proportional' punishment. Immanuel Kant proposed death. Other retributivists propose life in prison. Still others think life in prison is too harsh.

26. See Boonin, *The Problem of Punishment* (n 25); MJ Zimmerman, *The Immorality of Punishment* (Broadview Press 2011); and L Zaibert, *Rethinking Punishment* (Cambridge University Press 2018).

Second, Corrado argues that:

> If the legislature were controlled now and for all time by those who share Professor Caruso's attitudes, if we could hammer limits like his principle of least infringement into the Constitution and somehow or other ensure that it would be understood in the way he would have it understood, the worst possibilities that pre-emptive incapacitation and indefinite detention present to us might not play out.

While I understand Corrado's concern here, I do not understand why this is unique to my model. In the book, I repeatedly stress the importance of embracing my model in full, *with all its restrictions in place*, as well as fully funding the preventative and social justice components of my public health approach. It seems unfair to criticize the model by saying, 'Yeah, but consider what would happen if the model were only partially adopted, without its key protections and principles in place'. The same concern could be raised for just about any theory that proposes a holistic and comprehensive system.

This concern also applies to any reformist program that requires phasing in different reforms at different stages, since political will can, and often does, change before the full set of reforms are implemented. That's partly what happened with the government program of deinstitutionalization, which began in the 1960s as a way to improve the treatment of the mentally ill when widespread negligence and abuse was discovered in federal institutions. Deinstitutionalization was supposed to be *stage one* in a much larger program of reforms that included additional federal funding and capital investment in community mental health centres and services. Unfortunately, the government closed down their state-run institutions but failed to properly invest in the programs that were meant to replace them. As a result, today 3.5 million people considered severely mentally ill do not receive any psychiatric treatment[27] and our prisons have become de facto mental health institutions for which they are not properly designed. Sadly, this kind of thing happens all the time, but I do not think it provides good reason for embracing the status quo or giving up on reform.

Lastly, I will argue below in reply to Chloe Kennedy that while assessments of risk may retain a certain amount of indeterminacy and subjectivity, they still have a distinct advantage over the kinds of judgments required by retributivism and the Classical Model. If I am correct, then the potential for political abuse is even greater for those punitive theories that allow politicians and legislators to make judgments about desert, gravity and proportionality.

2 CHLOË KENNEDY

Chloë Kennedy's thoughtful and challenging comments focus on 'the substantial burden of persuasion that any alternative to retributive justice bears'.[28] She begins by noting that 'the intuitions that underlie the retributive ideal are powerful, widespread and tenacious'. Beyond this, she questions 'how fully the public health-quarantine model

27. Mental Illness Policy, 'About 60% of individuals with severe psychiatric disorders (3.5 million people) are receiving no treatment,' available at: <https://mentalillnesspolicy.org/consequences/percentage-mentally-ill-untreated.html>.
28. C Kennedy, 'Taking Responsibility for Criminal Responsibility: Comments on *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*' (2021) Journal of Legal Philosophy (this issue).

abandons the core intuitions of retributivism, including free will, and the extent to which it can discharge its own justificatory burden'. The majority of her comments focus on what she considers 'two similarities' between my account and the retributivist accounts I reject. First, she argues that there is 'some slippage' in the degree of control over our actions that I endorse. More specifically, she argues that in the second half of the book I rely on a 'degree of control over action' that I repudiate in the first half of the book. The effect, she argues, 'is that the conception of agency assumed in each account is not as different as might be expected'. Her second objection is that 'the foundational concepts in Caruso's account and the ones in the account he rejects are similar in that they are all politically and culturally contingent'.[29] Below I will address each of these concerns in turn, beginning with the latter.

2.1 Political and Cultural Contingency

Kennedy's concern about the contingency of concepts builds on my claim that judgments of gravity of wrong and proportionality are wide open to subjective and cultural biases and prejudices. While Kennedy agrees that those retributive concepts are culturally and politically contingent, she maintain that the same is true about judgments of 'harm and risk', which are foundational to my public health-quarantine model. She writes:

> The point I want to make … is that it is not clear that either harm or risk is any less 'tainted'. Even if we grant Caruso's assertion that retributive punishment is harder to justify than the public heath-quarantine responses, he admits that 'determining what counts as a significant threat to public health and safety is not always easy' and that 'the harm principle has been hotly debated'. Despite this, he concludes that harm might 'nevertheless serve as a useful guide in the domains of public health ethics and criminal justice'. With respect, I would suggest that this underplays the challenge.[30]

She goes on to conclude that: 'Just as questions about extra-legal morality are contested, so too are questions about what counts as harm and how far we should take account of remote and risked harms'.

In offering a reply, I would first acknowledge that Kennedy's concern is an important one and something I have, myself, struggled with. I acknowledge that assessments of risk and harm are difficult and typically involve some subjectivity. In the perfect world, this would not be the case. That said, I nevertheless contend that (a) a proper philosophical understanding of harm still has more constraints placed on it than judgments of gravity of wrong and proportionality. Furthermore, (b) risk assessment tools are constantly improving, and it's possible that one day we can develop methods that improve accuracy and, to whatever extent possible, control for and counterbalance human bias. The same, I contend, is not the case for judgments of wrong, blameworthiness and proportionality. Lastly, (c) even if it were the case that assessments of risk and harm would forever retain some indeterminacy, that would not itself mean that retributive punishment is *justified*. For the preceding reasons, I maintain that the public health-quarantine model has distinct advantages over retributivism. Let me expand on each of these points.

With regard to (a), I would argue that according to the public health-quarantine model – as well as JS Mill's *harm principle* and my own, more precise, *conflict*

29. ibid.
30. ibid.

resolution principle – the kinds of harms that would justify liberty-limiting restrictions are those that typically involve physical and bodily harm and threats to life, liberty and property. This is because the right of self-defence and defence of others is able to justify limiting one's liberty when that individual's actions seriously threaten another's life, liberty, property or physical well-being, but *not* in cases of offense, purely self-regarding acts or consensual private acts. That is, on my conception of harm, finding a particular action offensive does not constitute the kind of liberty-limiting harm at issue. If, for instance, someone was offended by a particular work of art at the Museum of Modern Art, that doesn't provide grounds for limiting the liberty of the artist to create it or the museum to exhibit it. Likewise, if one were left out of their grandmother's last will and testament, this may cause them some financial harm. But this is not enough to limit the liberty of the grandmother since the individual is not entitled to the money, nor are they harmed with respect to their rights of personal property, autonomy, or physical well-being. Furthermore, purely self-regarding acts and consensual acts between two adults, such as private sexual acts, would also be outside the scope of the right of self-defence since they do not harm any third parties, hence liberty should be preserved. Lastly, the right of self-defence would provide no grounds for limiting the liberty of individuals based purely on considerations of sexual orientation, religious belief, or freedom of conscience and expression since that would violate the harm principle and the conflict resolution principle.

This conception of harm places important constraints on the powers of the state that are simply not present in cases of retributivism, since deontological judgments of gravity and proportionality are too wide-open to cultural influence. To see this, simply examine the criminal codes and punitive practices of other cultures, times and places. Consider, for instance, the shifting attitudes about homosexuality. In the past, many countries viewed homosexuality as a grave wrong, punishable by the state. For instance, in the United Kingdom, under the Buggery Act of 1533, same-sex sexual activity was characterized as 'sinful' and was outlawed and punishable by death. The Offences Against the Person Act of 1861 removed the death penalty for homosexuality, but male homosexual acts remained illegal and were punishable by imprisonment. And while many liberal democracies have now abandoned such laws, there are unfortunately still many countries around the world where homosexuality is illegal and deemed a punishable offense. Examples like this reveal that what was once thought a grave wrong may no longer be. And what was once thought just and proportional punishment may later appeal cruel, inhumane and disproportionate. I give several additional examples in the book. Acknowledging this, however, raises a serious problem for the putative protective power of retributivism. How can the principle of proportionality prevent, in any reliable way, disproportionate punishment when there is no objective and impartial way to rank-order wrongs and determine what is to count as proportional punishment for each?

On the other hand, my conception of harm, grounded as it is in the defensive rights and constrained by the conflict resolution principle and harm principle, could never justify restricting the liberty of individuals when it comes to private consensual sexual acts. Furthermore, in the case of 'victimless crimes' where no one is harmed save the person engaging in the act, the harm principle would recommend decriminalization. The private use of marijuana likely constitutes such a case. But even if one disagrees, one thing is clear: Many of the low-level crimes we currently incarcerate people for (sometimes for many years) would be judged from the perspective of the public health-quarantine model as excessively punitive and unjustified. This is because significant weight should be given to protecting autonomy and individual liberty. Of course, such considerations can be outweighed when there is a significant threat of harm to public health and safety, but only in accordance with the conflict resolution principle and principle of least infringement. This generally means that incapacitation can only be justified when an individual's actions seriously threaten another's life, liberty, property or physical well-being.

At this point, Kennedy may object that, while these constraints on the concept of harm may have distinct advantages over judgments about desert, gravity and proportionality, it nonetheless remains difficult assessing when, and to what degree, an individual poses a serious enough threat to life, liberty, property or physical well-being. She correctly notes, for instance, that, 'parole boards tend to be risk averse and, unsurprisingly, their assessments are not free from bias or prejudice'.[31] I share this concern. That is why I recommend adopting additional policy solutions to help counterbalance these biases. The public health-quarantine model already maintains the principle of least infringement, which requires that we always seek alternatives to incapacitation wherever and whenever possible. But beyond this, I recommend establishing an upper limit of around 20 years in prison as a maximum penalty for the most dangerous of violent offenders, as Norway does, but to also allow for exceptional and incremental extensions when, say, a violent offender has not been rehabilitated and continues to pose a serious threat to public safety. Additionally, I recommend placing the burden of proof on the state to establish, at regular intervals, that the threat posed by an offender warrants continued incapacitation. If the state cannot satisfy the burden of proof with documentation, then the offender should be released. A review board comprised of psychologists and other professionals could make recommendations either to a judge or a parole board regarding whether continued confinement is necessary for public safety. And in such cases, they should propose appropriate treatment interventions designed to produce behavioural change leading to eventual release. Shifting the burden of proof to the state would therefore set the default in cases of parole at release, which the state would then need to overcome to justify continued incapacitation.

Lastly, I recommend that judges and parole boards adopt evidence-based violence risk assessments that help us identify historical and current risk factors as well as protective factors. An evidence-based violent risk assessment tool is a measure that has numerous meaningful violence risk factors that have been identified through research. Common evidence-based risk factors include substance use problems, psychopathic personality features, anger, impulsivity, antisocial peers, antisocial attitudes, a history of violence, young age at the first violent act, stress, treatment nonadherence, lack of social support and mental illness.[32] I acknowledge that such measures are not foolproof but they are the best tools we have and they are continually improving.

Consider, for instance, the Historical, Clinical, Risk-20 (HCR-20), which is a popular and commonly used structured professional judgment tool for assessing violence risk in adults.[33] The HCR-20 includes the following *historical risk factors* with guidelines on

31. ibid.
32. Emin Gharibian, 'Violence Risk Assessments: A Guide to Evaluating Violence Risk in Criminal Cases' Verdugo Psychological Associates (available online at: <https://verdugopsych.com/violence-risk-assessments-a-guide-to-evaluating-violence-risk-in-criminal-cases/>).
33. See KS Douglas, SD Hart, CD Webster and H Belfrage, *HCR-20V3: Assessing risk of violence – User guide* (Mental Health, Law, and Policy Institute, Simon Fraser University 2013). The HCR-20 has been subject to more than 200 empirical evaluations (see KS Douglas,

how to identify and code these risk factors: violence; other antisocial behaviour; relationships; employment; substance use; major mental disorders (e.g., psychotic disorders or serious mood disorders); personality disorders; traumatic experiences; violent attitudes; and response to treatment and supervision. *Recent clinical risk factors*, on the other hand, include: level of insight into their mental illness, violence risk, and need for treatment; violent ideation or intent; recent symptoms of a major mental disorder (e.g. psychotic disorders or serious mood disorders); affective, behavioural or cognitive instability; and compliance and response to treatment. Lastly, the *risk management scale* takes into consideration future problems with the following: professional services and plans; living situation; personal support; compliance and response to treatment; and stress management and coping skills. These risk and protective factors are taken into consideration to provide a risk formulation (low, moderate or high) regarding future violence risk, the risk of serious physical harm and risk for imminent violence. The formulation also allows for experts to identify risk management techniques that can be utilized to help mitigate and manage various risk factors.

To be clear, I'm not here endorsing HCR-20 or claiming that it's the best violent risk assessment tool available. I'm simply pointing out that evidence-based risk assessment tools are constantly being developed and improved, with the goal of increasing accuracy and reducing human subjectivity and bias. In fact, there are over 60 risk assessments measures currently being used in the United States,[34] and they have been shown to perform at a higher rate of accuracy than subjectivity of psychiatrists and parole boards.[35] And while these tools are not without problems and concerns of their own, they are constantly improving and it's possible that one day we can develop methods that improve accuracy even further and, to whatever extent possible, control for and counterbalance human bias. The same, I contend, is not the case for judgments of gravity of wrong, blameworthiness and proportionality. This is because (a) retributivism requires us to assess *two distinct* dimensions – judgments of gravity of wrong *and* judgments of proportionality – while violent risk assessment

C Shaffer, AJE Blanchard, LS Guy, K Reeves and J Weir, *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography* (Mental Health, Law, and Policy Institute, Simon Fraser University 2014). It has been translated into 20 languages, and adopted or evaluated in agencies within 35 countries. A recent large-scale survey by Singh et al. of 2,135 clinicians across 44 countries indicates that the HCR-20 was the most commonly used violence risk assessment instrument both in terms of assessing risk and creating risk management plans (see JP Singh, M Grann and S Fazel, 'A Comparative Study of Violence Risk Assessment Tools: A Systematic Review and Meta-Regression Analysis of 68 Studies Involving 25,980 Participants' (2011) 31 Clinical Psychology Review 499).

34. See AM Barry-Jester, B Casselman, D Goldstein, M Conlen, R Fischer-Baum and A Rossback, *Should Prison Sentences be Based on Crimes That Haven't Been Committed Yet?* (2015) available at: <https://fivethirtyeight.com/features/prison-reform-risk-assessment/> (accessed 10 January 2020); and B Casselman and D Goldstein, *The New Science of Sentencing* (2015) available at: <www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing> (accessed 14 August 2019).

35. See WM Grove, DH Zald, BS Lebow, BE Snitz and C Nelson, 'Clinical versus Mechanical Prediction: A Meta-Analysis' (2000) 12 Psychol. Assess. 19–30; S Ægisdóttir, MJ White, PM Spengler, AS Maugherman, LA Anderson, RS Cook, et al., 'The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical versus statistical prediction' (2006) 34 The Counseling Psychologist 341; CC Spohn, *How Do Judges Decide? The Search for Fairness and Justice in Punishment*, 2nd Edition (Sage Publications 2008); J Dressel and H Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4 Sci. Adv. 5580.

measures need only assess one dimension. Retributivism therefore allows for cultural and political bias to enter the equation at twice the number of places. (b) We have a good (and improving) understanding of the social, psychological and neuroscientific determinants of violence, while there is no equivalent evidence-based approach to understanding gravity, desert and proportionality. This is made even worse by the fact that (c) judgments about gravity, desert and proportionality are value laden and normative in nature, whereas assessments of risk are more descriptive in nature and amenable to scientific study and understanding. Lastly, (c) since my conception of harm is grounded in the defensive rights and constrained by the conflict resolution principle and harm principle, it better protects the liberty of individuals in those cases that do not involve a significant risk of harm to another's life, liberty, property or physical well-being. For these reasons, I maintain that the public health-quarantine model has a number of distinct advantages over retributivism.

My final reply to Kennedy would be to note that even if it were the case that assessments of risk and harm would forever retain some indeterminacy, that would not itself mean that retributive punishment is *justified*. Since retributive punishment requires the kind of free will needed for basic desert moral responsibility, free will skepticism implies that retributive punishment lacks justification. To the extent, then, that my arguments against retributivism succeed, I contend that we should reject retributivism in favour of justified legal practices. Kennedy's objection therefore fails to establish that my model is subjective to the same extent as retributivism or that it should be abandoned because risk assessment methods are not a hundred percent reliable. Additionally, until and unless retributivism can overcome my skeptical and epistemic arguments, something Kennedy does not attempt to do, we should conclude that retributive legal punishment is unjustified and the intentional harms it causes are prima facie seriously wrong.

2.2 The Scope of Agency

Moving on to Kennedy's second main objection, she argues that there's 'slippage' in the degree of control over our actions that I endorse. She argues: 'Caruso is unable to support the forward-looking forms of responsibility he wants to endorse – rehabilitation, therapy, acknowledgement of wrongdoing – without altering his position with respect to control over action across the course of his book'. She points to the following potential points of conflict: (a) First, she argues that 'Caruso tells us that free will skeptics need not deny 'the causal efficacy of our choices and intentions' so long as these are understood to be constrained by the circumstances. It is not clear how this claim differs from that of the retributivist who wants to acknowledge that freedom comes in degrees – a claim Caruso apparently rejects'. She then goes on to (b) challenge my claim that rational deliberation is compatible with determinism since, I argue, it only requires *epistemic openness* and a *deliberative-efficacy condition*. In reply, she argues that 'this is not how we tend to deliberate – we tend to believe our decisions to act are likely, maybe even very likely, to result in our taking that action'. Furthermore, (c) she contends that, 'the kind of control over one's future beliefs and actions that therapy and rehabilitation would seem to require, i.e., actual control, is not clearly available on Caruso's account'. Lastly, (d) she questions 'why a freedom that could potentially ground forward-looking responsibility should not ground backward-looking responsibility as well'.

Before replying to the specifics of each of these concerns, let me begin by reiterating that as a free will skeptic I do not deny that there are important differences between agents who have the kind of control compatibilists have identified. Such distinctions are undeniable. A normal adult who is responsive to reasons, for instance, differs in significant ways from one who is suffering from psychopathy, Alzheimer's or severe mental illness. And an agent whose effective desire conforms appropriately to his/her 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 – i.e., the kind of responsibility that would make us truly deserving of blame and praise, punishment and reward in a purely backward-looking sense.

Keeping that in mind, my response to (a) should be obvious. Acknowledging, as I do, the different levels of control an agent can possess, as well as the causal efficacy of our choices and intentions, *differs* from the 'retributivist who wants to acknowledge that freedom comes in degrees' *in that* the retributivism assumes that such control is equivalent to free will and sufficient (along, perhaps, with some additional conditions) for basic desert moral responsibility. I deny that that is the case. Kennedy seems to assume that acknowledging the 'the causal efficacy of our choices' is somehow enough to ground the retributive assumption that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done in a purely backward-looking, non-consequentialist sense. I contend that that is not the case. If Kennedy wants to defend that assumption, she needs to overcome not just one, but a series of arguments, including the manipulation argument, the luck pincer and the epistemic argument.

With regard to (b), I have a similar reply. Here Kennedy is challenging my claim that rational deliberation is compatible with determinism since it only requires *epistemic openness* and a *deliberative-efficacy condition*. If she wishes to reject that claim, then she *must also* conclude that compatibilism is unable to preserve rational deliberation – since deliberation-compatibilism is a view I share in common with compatibilists. *Deliberation-Compatibilism* maintains that *S*'s deliberating and being rational is compatible with *S*'s believing that their actions are causally determined by antecedent conditions beyond their control. As a compatibilist, Kennedy does not want to reject that claim – hence, we are *both* deliberation-compatibilists. The difference between us, however, is that I further argue that the notion of *sourcehood* relevant to rational deliberation differs from that involved in free will.[36] Furthermore, I do not deny, as Kennedy seems to suggest, that 'our decisions to act … result in our taking that action'. In fact, I argue at length elsewhere that for an agent to exercise the kind of causal influence relevant to rational deliberation, the causal chain must pass through their agential structures.[37] I call this the *deliberative causal influence* (DCI) principle and define it more accurately as follows:

> **DCI:** For an agent *S* to exercise the kind of causal influence relevant to rational deliberation, the causal chain must pass through *S*'s agential structures and the resulting action, whatever it ends up being, be the result of *S*'s decision – where 'be the result of' is understood causally.

36. See GD Caruso, 'On the Compatibility of Rational Deliberation and Determinism: Why Deterministic Manipulation Is Not a Counterexample' (2021) 71 The Philosophical Quarterly 524.
37. ibid.

My view is that rational deliberation requires only the minimal kind of control specified in (DCI), and perhaps some other non-controversial conditions, whereas the control in action required for free will requires a more robust notion of control. Once we separate these two notions and see that rational deliberation only requires belief in the weaker notion of control, Kennedy's concern goes away.

Moving on to (c), Kennedy contends that, 'the kind of control over one's future beliefs and actions that therapy and rehabilitation would seem to require, i.e., actual control, is not clearly available on Caruso's account'.[38] Here I'm simply not sure what Kennedy means by 'actual control'. If she means something like libertarian control, then clearly we disagree. But if she means only that rehabilitation can help restore the kind of control needed for agents to be responsive to reasons, approve of their effective desires, have deliberative causal influence over their decisions, and the like, then I agree with Kennedy that such control is important when it comes to therapy and rehabilitation. But at the cost of sounding like a broken record, I do not deny such control. Instead, I argue that such control 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.

Lastly, (d) Kennedy questions 'why a freedom that could potentially ground forward-looking responsibility should not ground backward-looking responsibility as well'.[39] The answer, I contend, again lies in *basic desert* moral responsibility. On the forward-looking account of moral responsibility Derk Pereboom and I endorse,[40] when we encounter apparently immoral behaviour, it is perfectly legitimate to engage in moral protest and ask the agent, 'Why did you decide to do that?' or 'Do you think it was the right thing to do?' If the reasons given in response to such questions are morally unsatisfactory, we regard it as justified to invite the agent to evaluate critically what their actions indicate about their intentions and character, to demand an apology, or request reform. We maintain that engaging in such interactions is reasonable in light of three forward-looking considerations, the first being the right of those harmed or threatened to protect themselves from immoral behaviour and its consequences. Second, we might have a stake in reconciliation with the wrongdoer, and calling them to account in this way can function as a step toward realizing this objective. Lastly, on both a personal and societal level we have an interest in the moral formation of the wrongdoer, and the address described functions as a stage in that process. On this forward-looking account of moral responsibility, then, moral protest and exchange is grounded, *not in basic desert*, but in *three non-desert-invoking desiderata*: future protection, future reconciliation and future moral formation. Since this account in no way appeals to basic desert or its controversial backward-looking practices, we maintain that it is perfectly consistent with the rejection of free will. Kennedy has given me no reason for doubting this.

38. Kennedy, 'Taking responsibility for criminal responsibility: Comments on *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*'.
39. ibid.
40. See D Pereboom, *Free Will, Agency, and Meaning in Life* (Oxford University Press 2014); and Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*.

3 REPLIES TO KATRINA SIFFERD

Katrina Sifferd has been a friend and worthy adversary for many years. My exchanges with her have always been rewarding and fruitful. Her critical comments here are no exception. She begins by agreeing with me that our current system of criminal justice needs major reform – since ‘sentences are often disproportionate and overly punitive’ and ‘prisons serve to encourage recidivism’. Nevertheless, she thinks my view ‘throws the baby out with the bathwater’. Instead, she argues that, ‘it might be possible to reform the US criminal justice system to look more like the Nordic systems Caruso lauds in the book’,[41] while still preserving a form of ‘weak’ retributivism – one that appeals to a mixed theory of justification. Below, I begin by addressing some comments Sifferd makes about my free will skepticism and my reasons for rejecting retributivism. I then examine Sifferd’s own preferred mixed theory of justification and explain why I reject it. I conclude by addressing Sifferd’s concern that my public health-quarantine model ‘encourages the state to rehabilitate persons in ways that may infringe upon their rights’.

3.1 Free Will Skepticism

To begin, Sifferd takes issue with my free will skepticism and maintains that, contra my arguments for hard-incompatibilism and hard-luck, a reasons-responsive compatibilist theory of responsibility remains justified – one that maintains that reasons-responsiveness and the capacity to be affected by and respond to the demands of the law ‘are enough to ground basic desert responsibility within the context of a mixed justification of criminal law and punishment’. Putting aside Sifferd’s mixed justification of punishment for a moment, I do not think Sifferd takes seriously enough the arguments against reasons-responsive accounts of moral responsibility – in particular, the manipulation argument and the luck pincer. In response to the manipulation argument, for instance, Sifferd provides only the following short reply:

> I think Deery and Nahmias have an adequate reply to manipulation arguments except in some rare and far-fetched/philosophically contrived cases. On their view, an agent may be seen as the causal source of their actions where no other variable has a stronger causal relationship than those within the agent’s control. This seems right.[42]

This comprises the full extent of Sifferd’s reply to the manipulation argument and I find it insufficient for a number of reasons. Most importantly, it overlooks the long and detailed response to Deery and Nahmias[43] I provide in Section 2.2.2.4 of the book as well as the objections raised by Tierney and Glick[44] and Pereboom and McKenna.[45]

41. K Sifferd, ‘Why Not “Weak” Retributivism?’ (2021) Journal of Legal Philosophy (this issue).
42. ibid.
43. O Deery and E Nahmias, ‘Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood’ (2017) 174 Philosophical Studies 1255.
44. H Tierney and D Glick, ‘Desperately Seeking Sourcehood’ (2000) 177 Philosophical Studies 953.
45. D Pereboom and M McKenna, ‘Manipulation Arguments against Compatibilism’, in D Kay Nelkin and D Pereboom (eds), *Oxford Handbook of Moral Responsibility* (Oxford University Press 2021).

The central idea of Deery and Nahmias's account is that causal variables that can result in an outcome by more than one means – that is, in response to a wider range of changes to the background conditions – bear stronger causal invariable relations to their outcome variables than variables that cannot, or than variables that only result in the outcome across a narrower range of changes to the background conditions.[46] In the case of natural determinism, Deery and Nahmias argue that the output of deliberative activity within the agent's compatibilist agential structure bears a stronger causal invariance relation to their actions than any other variable – that is, the output of the compatibilist agential structure will cause the action across a wider range of circumstances. In manipulation cases, on the other hand, Deery and Nahmias argue that this is not the case. Instead, they contend, the strongest causal invariable relation exists between the decision of the manipulator(s) to manipulate the agent. In this way, Deery and Nahmias claim to have isolated a relevant difference between naturally determined agents and manipulated agents.

While this reply to manipulation cases may be a slight improvement over previous replies, the account still faces significant problems. First and foremost, it does not take into account the fact that both the nature of manipulation and the nature of the manipulated agent can be made to vary, with the result that the two can be made to match in degrees and kinds of invariance. For Deery and Nahmias, in manipulation cases it is important that the external manipulator(s) both *intend* for a particular action or decision to occur and are able to *ensure* that it occurs. It is the combination of these two factors that allow Deery and Nahmias to claim that the strongest causal invariance relation exists between the manipulators' intention/decision and the agent's action. The problem with this reply, however, is that manipulation cases need not assume such invariance, nor must they require manipulators to be able to 'ensure' specific outcomes. In fact, manipulation cases can be devised where the manipulation is neither intended not ensured. My *Brain-Implant Malfunction Case* is just such a counterexample:[47]

> Imagine that Plum has a device implanted in his brain for medical purposes (e.g., to control seizures or trembling due to Parkinson's disease), not for the purpose of manipulation. When operating normally, the device does not affect Plum's reasoning. One day the device malfunctions and ends up triggering in Plum a strong egoistic reasoning process that deterministically results in his decision to kill White. And it does so in a way that satisfies all the prominent compatibilist conditions (e.g., Plum is moderately reasons-responsive, approves of his decision to kill White, is able to grasp, apply, and regulate his actions by moral reasons, etc.). Here we have a case of accidental manipulation due to a malfunctioning device implanted in Plum's brain.

Here the manipulated agent is determined to commit a criminal action *without external agents acting as intentional manipulators*. And it would not aid the soft-liner to reply that since the device was intentionally implanted the case fails, since the fact that the device was intentionally implanted (for medical purposes) in no way undermines the main philosophical point: the malfunctioning of the device, and the subsequent manipulation, was *not the result of external agents acting as intentional manipulators*. Hence, the presence of intentional manipulation by external agents cannot be the relevant difference soft-line theorists are after. Furthermore, the responsibility-undermining manipulation would *not* provide the strongest causal invariance relation

46. Deery and Nahmias, 'Defeating Manipulation Arguments' (n 43) 1263.
47. Caruso (n 1) 85.

to the agent's action, since small changes in the background conditions would not result in the same output. Since it is *false*, then, that in this case there are very few changes in background circumstance that could break the causal connection between the malfunctioning of the brain-implant and the manipulated action, my Brain-Implant Malfunction Case provides a counterexample to Deery and Nahmias's analysis.

Sifferd does not address this or other recent replies to Deery and Nahmias.[48] She also unceremoniously dismisses the other main argument against compatibilism defended in the book: Neil Levy's *luck pincer*.[49] At the heart of the argument is the following dilemma: either actions are subject to *present luck* (luck around the time of action), or they are subject to *constitutive luck* (luck that causes relevant properties of agents, such as their desires, beliefs, values and circumstances), or both. Either way, it is argued, luck undermines moral responsibility since it undermines responsibility-level control. Robert Hartman[50] summarizes this argument as follows:

> *Universal Luck Premise*: Every morally significant act is either constitutively lucky, presently lucky, or both.
>
> *Responsibility Negation Premise*: Constitutive and present luck each negate moral responsibility.
>
> *Conclusion*: An agent is not morally responsible for any morally significant acts.

While I consider the arguments for hard indeterminism to be sufficient on their own for concluding that agents are never morally responsible in the basic desert sense, I think the same for the luck pincer. That is, I consider it a second, independent argument for free will skepticism. Hence, defenders of free will, as well as retributivists who ground their justification for punishment in the notion of *just deserts*, must overcome both sets of arguments. It's not enough to argue that one of these routes to free will skepticism fails, since if the right hand doesn't get you, the left hand will.

Consider, for instance, the problem constitutive luck raises for the compatibilist. Since our genes, parents, peers and other environmental influences all contribute to making us who we are, and since we have no control over these, it seems that who we are is largely a matter of luck. And since how we act is partly a function of who we are, the existence of constitutive luck entails that what actions we perform depends on luck. A compatibilist could respond, as they often do, that as long as an agent *takes responsibility* for her endowments, dispositions and values, over time she will *become* morally responsible for them. The problem with this reply, however, is that the series of actions through which agents shape and modify their endowments, dispositions and values are *themselves* subject to luck – and as Levy puts it, 'we cannot undo the effects of luck with more luck.'[51] Hence, the very actions to which compatibilists point, the actions whereby agents take responsibility for their endowments, either *express* that endowment (when they are explained by constitutive luck) or reflect the agent's present luck, or both. Hence, the luck pincer.

48. For a counterexample similar to my own, see Pereboom and McKenna, 'Manipulation Arguments against Compatibilism' (n 45). For a completely different type of objection to Deery and Nahmias, see Tierney and Glick, 'Desperately Seeking Sourcehood' (n 44).
49. Neil Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility* (Oxford University Press 2011).
50. Robert Hartman, *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness* (Routledge 2017).
51. Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility* (n 49) 244.

In response to this argument, Sifferd simply points to the fact that 'several … prominent philosophers' disagree with its conclusion. She cites in particular Robert Hartman.[52] This appeal to Hartman, however, is unconvincing (at least without further argument) since I dedicate an entire section of the book to responding to Hartman's objections, concluding that each can be satisfactorily dealt with and as a result we should conclude that the skeptical view remains the most justified position to adopt since the pervasiveness of luck undermines free will and basic desert moral responsibility.[53] Without further argument, then, I see no reason for thinking that a reasons-responsive compatibilist account can overcome the arguments I present in chapter 2 of the book.

3.2 Sifferd's Mixed Theory

Moving on, then, to Sifferd's own positive proposals, she defends a form of 'weak' retributivism according to which, 'desert is necessary, but not sufficient, for punishment'. She writes: 'According to my preferred theory, persons who commit crimes forfeit their right not be punished; however, there are other important justifications for criminal punishment based upon its effects; and if these justifications did not hold, then institutional punishment by the state may not be justified'. She also defends a forward-looking account of backward-looking moral responsibility similar to that of Manuel Vargas, Daniel Dennett and Victoria McGeer.[54] On this account, the practices involved in holding agents morally responsible in the *desert* sense are justified by 'pointing to these practices' forward-looking effects on morally sensitive agency and reasons-responsiveness'. More specifically, 'the reactive attitudes, including moral blame, are justified [on forward-looking grounds] if they appropriate practices to violations of society's norms because these norms are crucial to the responsibility practices that enhance us as moral agents'. For Sifferd, then, 'the law may be justified instrumentally because it provides important support to responsible agency by encouraging diachronic management of decision-making, dispositions and environment such that an agent is more likely to meet the law's expectations'.

Before explaining why I reject Sifferd's proposals, let me first note that Sifferd's weak retributivism and her mixed account of moral responsibility are *two distinct theses* and should be judged independently of each other. One could, for instance, embrace weak retributivism while subscribing to a purely backward-looking account of desert – i.e., one could reject Sifferd's mixed account of moral responsibility and still be a weak retributivist. On the other hand, one could appeal to an instrumentalist account of desert in an attempt to justify *more than* just weak retributivism. There are, for instance, mixed views and two-tiered systems that appeal to both forward- and backward-looking considerations but nonetheless embrace more robust forms of retributivism. I think it best, then, to evaluate Sifferd's proposals separately, beginning with her weak retributivism.

52. Hartman, *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness* (n 50).
53. Caruso (n 1) Sect. 2.3.2.
54. M Vargas, *Building Better Beings: A Theory of Moral Responsibility* (Oxford University Press 2013); D Dennett, *Elbow Room: Varieties of Free Will Worth Wanting* (MIT Press 1984); D Dennett and GD Caruso, *Just Deserts: Debating Free Will* (Polity Books 2021); V McGeer, 'Building a Better Theory of Responsibility' (2015) 172 Philosophical Studies 2635.

Sifferd is correct that I do not dedicate much space to discussing weak retributivism in the book, choosing instead to focus on *moderate* and *strong* retributivism. There are, however, several good reasons for this. First, most leading retributivists defend either moderate or strong versions of retributivism[55] and it was my desire in the book to address the dominant view, not a subordinate view held by few. Second, the weight the criminal law gives desert and the way retributivism is practically implemented in the law (especially in the United States) indicate that the desert of offenders is typically seen as sufficient for punishment. The revised Model Penal Code, for instance, makes this point rather clear.[56]

Third, weak retributivism is considered by many retributivists to be 'too weak to guide the criminal law' and as amounting to nothing more than 'desert-free consequentialism side-constrained by negative desert'.[57] In fact, some theorists simply define retributivism in a way that excludes weak retributivism from consideration altogether. David Boonin, for example, defines retributivism as the claim that 'committing an offense in the past is *sufficient* to justify punishment now, whether or not this will produce any beneficial consequences in the future'.[58] Retributivist Mitchell Berman maintains that the 'core retributivist thesis' is that:

> [t]he goodness or rightness of satisfying a wrongdoer's negative desert morally justifies [i.e., is sufficient for] the infliction of criminal punishment, without regard for any further good consequences that might be realized as a contingent result of satisfying the wrongdoer's desert.[59]

And Alec Walen in his *Stanford Encyclopedia of Philosophy* entry on 'Retributive Justice'[60] defines retributivism as committed to the following three principles: (1) that those who commit certain kinds of wrongful acts, paradigmatically serious crimes, morally deserve a proportionate punishment; (2) that it is intrinsically morally good – good without reference to any other goods that might arise – if some legitimate punisher gives them the punishment they deserve; and (3) that it is morally

55. See, for example, M Moore, *Placing Blame* (Oxford University Press 1997); M Moore, 'The Moral Worth of Retribution', in JG Murphy (ed), *Punishment and Rehabilitation* (Wadsworth Publishing Company); M Moore, *Act and Crime: The Philosophy of Action and Its Implications for Criminal Law* (Oxford University Press 1993); S Kershnar, 'A Defense of Retribution' (2000) 14 International Journal of Applied Philosophy 97; S Kershnar, *Desert, Retribution, and Torture* (University Press of America 2001); D Husak, 'Holistic Retributivism' (2000) 88 California Law Review 991; M Berman, 'Punishment and Justification' (2008) 18 Ethics 258; M Berman, 'Two Kinds of Retributivism', in RA Duff and S Green (eds), *Philosophical Foundations of Criminal Law* (Oxford University Press 2011); A Von Hirsch, *Doing Justice: The Choice of Punishment* (Northeastern University Press 1976); A Von Hirsch, 'The "Desert" Model for Sentencing: Its Influence, Prospects, and Alternatives' (2007) 74 Social Research 413; A Von Hirsch, *Deserving Criminal Sentences: An Overview* (Hart Publishing 2017); L Alexander, 'You Got What You Deserved' (2013) 7 Criminal Law and Philosophy 309; L Alexander, K Ferzan and S Morse, *Crime and Culpability: A Theory of Criminal Law* (Cambridge University Press 2009).
56. See Caruso (n 1) 7–8.
57. Alexander, Ferzan and Morse, *Crime and Culpability: A Theory of Criminal Law* (n 55) 7.
58. Boonin, *The Problem of Punishment* (n 25) 86 (emphasis added).
59. M Berman, 'Modest Retributivism', in K Kessler and SJ Morse (eds), *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore* (Oxford University Press 2016).
60. A Walen, 'Retributive justice' *Stanford Encyclopedia of Philosophy*' (2014) available at: <https://plato.stanford.edu/entries/justice-retributive/>.

impermissible to intentionally punish the innocent or to inflict disproportionately large punishments on wrongdoers. Given that this is how most philosophers understand retributivism, I thought it best to target the claim that the desert of offenders provides sufficient grounds for punishment and that we are therefore justified in sometimes punishing wrongdoers for no purpose other than to see the guilty get what they deserve.

Fourth, since weak retributivism is unable to justify legal punishment on its own, it requires an appeal to additional, usually consequentialist, justifications. Since I dedicate an entire chapter to independently evaluating and arguing against these additional non-retributive justifications of punishment, weak retributivism is already sufficiently, though indirectly, dealt with. Defenders of weak retributivism would not only need to defend the notion of desert but they would also need to overcome the objections I and others have presented against consequentialist and other non-retributive justifications.

Lastly, and perhaps most important given Sifferd's own reformist proclivities, I don't see much advantage in defending a limited form of retributivism that preserves only the notion of *negative desert*.[61] Mario DeCaro, for instance, defends such an account and maintains that we should abandon the notion of *positive desert* for the kind of skeptical reasons I defend in the book.[62] DeCaro accepts, for instance, the basic conclusion of the Epistemic Argument and thinks doubts about the existence of free will are sufficient to reject positive retributivism. Nevertheless, he proposes that we maintain the protections of negative desert. While I am sympathetic to this position and think it is a marked improvement over traditional retributivism, it is important to recognize that it has much more in common with my free will skepticism than it does with most traditional forms of retributivism. Such a view would likely be unacceptable to most retributivists. Furthermore, I think the preservation of negative desert is unnecessary since there are other resources the skeptic can appeal to that provide the same level of protection against punishing innocent people that negative desert does. For instance, on my account, competent and reasons-responsive agents would need to be *causally* responsible for committing a particular criminal offense before incapacitation could be justified for all the reasons explained in my reply to Corrado.

Turning now to Sifferd's mixed account of moral responsibility, I would first reiterate that the kind of moral responsibility I deny – and the kind I contend is of central philosophical and practical importance when it comes to our interpersonal, moral and legal desert-based practices, attitudes and judgments – is *basic desert*

61. *Negative desert* can be contrasted with *positive desert*, which has to do with an agent deserving praise or reward for good actions. It's important to note that there is another conception of 'negative desert' that is widespread in the literature. This latter notion refers to the negative component of the retributivist thesis. As Walen examples, 'Retributivism … involves both positive and negative desert claims. The positive desert claim holds that wrongdoers morally deserve punishment for their wrongful acts.' On the other hand, '[t]his positive desert claim is complemented by a negative one: Those who have done no wrong may not be punished. This prohibits both punishing those not guilty of wrongdoing (who deserve no punishment), and punishing the guilty more than they deserve (i.e., inflicting disproportional punishment)' (Walen, 'Retributive justice' (n 60) sec. 3.1). Having two different notions of negative desert can potentially be confusing, but I will try my best to make clear which conception is at play in different contexts.
62. M DeCaro, 'The Indispensability of the Manifest Image' (2020) 46 Philosophy and Social Criticism 113.

moral responsibility. It is important to distinguish *basic desert* from those accounts of *non-basic desert* that attempt to justify the *moral responsibility system* or *system of desert* on consequentialist grounds. On such forward-looking accounts of moral responsibility, practice-level justifications for blame and punishment invoke considerations of desert, while that desert is not basic because at a higher level the practice is justified by good anticipated consequences, such as deterrence of wrongdoing and moral formation of wrongdoers. According to such accounts, our practice of holding agents morally responsible in this non-basic desert sense should be retained because doing so would have the best overall consequences relative to alternative practices.

Against such views I have offered three main replies. First, I have argued that penalties and rewards justified by anticipated consequences do not really qualify as genuinely *deserved* since on such a view they ultimately function as incentives.[63] Second, it remains an open question whether preserving the moral responsibility system and 'system of desert' has the forward-looking benefits Sifferd, Vargas, Dennett and McGeer maintain. The notion of *just desert*, for instance, is too often used to justify punitive excess in criminal justice, to encourage treating people in severe and demeaning ways, and to excuse and perpetuate social and economic inequalities. Additionally, resentment, indignation, moral anger and blame are often counterproductive on the interpersonal level when it comes to the goals of safety, moral formation and reconciliation. In fact, there is growing empirical evidence that free will beliefs are themselves motivated by the desire to punish others and to justify holding them morally responsible.[64] Researchers have also found that free will beliefs correlate with increased punitiveness.[65] If this is correct, then leaving the concepts of free will and desert unchallenged increases the likelihood that our practices and policies will remain focused on individual responsibility and punitive responses to wrongdoing. This, in turn, will stand in the way of adopting the kind of reform I recommend (and Sifferd seems open to), which shifts the focus to prevention, rehabilitation and addressing the social determinants of criminal behaviour.

Lastly, such accounts of moral responsibility do not allow us to understand the substantive disputes that drive the free will debate. Some philosophers, for instance, identify themselves as compatibilists because they hold that some *non-basic-desert* notion of moral responsibility, often one they regard as sufficient for the moral life, is compatible with determinism.[66] But if 'compatibilism' is defined so that such a position turns out to be compatibilist, virtually everyone in the debate stands to be a compatibilist. Daniel Dennett, for instance, specifies that his compatibilist notion of free will can 'play all of the valuable roles free will has been traditionally invoked

63. See Dennett and Caruso, *Just Deserts: Debating Free Will* (n 54).
64. See, e.g., CJ Clark, JB Luguri, PH Ditto, J Knobe, AF Shariff and RF Baumeister, 'Free to Punish: A Motivated Account of Free Will' (2014) 106 Journal of Personal and Social Psychology 501.
65. JM Carey and DL Paulhus, 'Worldview Implication of Believing in Free Will and/or Determinism: Politics, Morality, and Punitiveness' (2013) 81 Journal of Personality 131; T Nadelhoffer and DG Tocchetto, 'The Potential Dark Side of Believing in Free Will (and Related Concepts): Some Preliminary Findings', in GD Caruso (ed), *Exploring the Illusion of Free Will and Moral Responsibility* (Lexington Books 2013); AF Shariff, JD Greene, JC Karremans, J Luguri, CJ Clark, JW Schooler, RF Baumesiter and KD Vohs, 'Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution' (2014) Psychological Science.
66. Dennett, *Elbow Room* (n 54); D. Dennett, *Freedom Evolves* (Penguin Books 2004); Vargas, *Building Better Beings: A Theory of Moral Responsibility* (n 54).

to play'.[67] Stephen Morris, however, aptly complains that Dennett 'has defined the conceptions of "free will" and "moral responsibility" in such a way to eliminate any substantive differences between the 'compatibilist' position he defends and the hard determinist position that philosophers typically understand as being substantively different from compatibilism'. I stress this same point repeatedly in my debate with Dennett.[68] Since free will skeptics can defend a forward-looking account of moral responsibility – one grounded not in basic desert, but in *three non-desert-invoking desiderata*: future protection, future reconciliation and future moral formation – the compatibilist must have something more controversial in mind, otherwise the free will debate becomes purely semantic with no substantive disagreement between the skeptic and compatibilist.

By way of remedy, Pereboom and I have both contended that the compatibilist/incompatibilist terminology should reflect the lines of division in the debate, and thus it would be best to use the term 'compatibilism' to designate a view in which an agent's having the sort of free will required for moral responsibility in the basic desert sense is compatible with her actions being causally determined by factors beyond her control.[69] Straying from this characterization threatens to make the terminology valueless for characterizing core opposing positions in the debate, and to give rise to merely verbal disputes.

3.3 Rehabilitation and Restraint in Public Quarantine

Sifferd's final objection is that a system of criminal justice that preserves, on instrumentalist grounds, moral responsibility and weak retributivism is 'more likely to respect the dignity and autonomy of offenders than Caruso's system of public quarantine'. While I take this concern seriously, I maintain that it is wrong to think that free will skeptics cannot preserve respect for human dignity or protect against harsh and inhumane treatment without the notion of basic desert.

In response, I offer three main replies. First, I contend that the retributive principle of proportionality, in actual practice, does not guarantee respect for persons any better than the alternatives. This is because measurements of gravity and proportionality are hypersensitive to cultural biases, prejudices and power relations, and there remain serious problems for both cardinal and ordinal ways of measuring proportionality. Second, the public health-quarantine model, has a non-desert-based principle of proportionality of its own, one which is capable of securing respect for persons and protecting innocent people from being used as a means to an end. Lastly, respecting human dignity, I argue, is not about giving wrongdoers their just deserts. Rather, it demands that the capabilities and well-being of wrongdoers be taken into consideration, that we avoid punishments that dehumanize and disenfranchise individuals, and that we do everything we can to rehabilitate and reintegrate offenders back into the

67. S Morris, 'The Impact of Neuroscience on the Free Will Debate' (2009) 9 Florida Philosophical Review 69.
68. Dennett and Caruso, *Just Deserts: Debating Free Will* (n 54).
69. See, e.g., Pereboom, *Living Without Free Will* (Cambridge University Press 2011); Pereboom, *Free Will, Agency, and Meaning in Life* (n 40); GD Caruso and S Morris, 'Compatibilism and Retributive Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance' (2017) 82 Erkenntnis 837; GD Caruso and D Pereboom, *Moral Responsibility Skepticism* (Cambridge University Press, forthcoming).

community. The public health-quarantine model does a better job at respecting human dignity in this sense than does retributivism.

Consider, for instance, the fact that while the public health-quarantine model rejects the retributivist principle of proportionality, it has a proportionality principle of its own. This is because the public health-quarantine model maintains that legal sanctions should be proportionate to the danger posed by an individual, and any sanctions that exceed this upper bound will be unjustified. This is coupled with the *principle of least infringement* and the *conflict resolution principle*. These principles set strict limits on how individuals can and should be treated. They prohibit, for instance, the incapacitation of innocent people who pose no threat to society.[70] On the public health-quarantine model, the justification for incapacitation should not be understood in a strict consequentialist theoretical context. Rather, the model justifies incapacitation on the ground of the right to self-defence and defence of others. That right does not extend to people who are non-threats. It would therefore be wrong to incapacitate someone who is innocent since they are not a serious threat to society.

The right of self-defence can only justify limiting one's liberty when that individual's actions seriously threaten another's life, liberty, property, or physical well-being. Since innocent people do not pose such a threat, it would be a violation of the conflict resolution principle to incapacitate them. Hence, the public health-quarantine model, just like retributivism, is able to prohibit the punishment/incapacitation of innocent people. The principle of least infringement would also prohibit legal sanctions that exceed what is needed to protect public health and safety. As a result, it would oppose using individuals simply as a means to deter others by ratcheting up various punitive responses to crime, as three strikes laws did. For this reason, the public health-quarantine model also has certain advantages over consequentialist theories of punishment. I therefore contend that the public health-quarantine model is able to provide the kinds of protections needed to protect innocent people from legal sanctions and prohibit excessively punitive forms of punishment.

I also maintain that the public health-quarantine model does a better job respecting human dignity than does retributivism. Consider, for instance, the following hypothetical scenario used by Shariff and colleagues. The fictional case involves an offender who beat a man to death but after serving two years in prison was nearly 100 percent effectively rehabilitated. The case further stipulates that 'the prosecution and defence had agreed that the rehabilitation would prevent recidivism and that any further detention after rehabilitation would offer no additional deterrence of other potential criminals'.[71] While this is clearly a contrived scenario, we can use it here as a test case. On my model, it would be unjust to continue to incapacitate this individual. Retributivists, on the other hand, will generally feel that this person deserves more than two years in prison (though they will likely disagree on how much more). I wonder which of these views better respects human dignity. I contend that punishing someone who is no longer a threat to society, and in a way that exceeds effectiveness, is not the proper way to respect human dignity. Instead, I maintain that the public health-quarantine model actually respects human dignity more since it specifies that (a) individuals who are not a serious threat to society should not be incapacitated, (b) no one should be incapacitated longer than is absolutely necessary (where this is determined by the continued threat the individual poses to society), and (c) when it is necessary to

70. See again my reply to Corrado on this point.
71. Shariff et al., 'Free Will and Punishment: A Mechanistic View of Human Nature Reduces Retribution' (n 65) 4.

incapacitate an individual, we must do so in a way that treats them humanely, with respect and dignity, and with rehabilitation as our goal.

Furthermore, the *capabilities approach* to social justice, which I use to ground my public health framework, demands that we take into consideration the well-being of wrongdoers, that we avoid punishments that dehumanize and disenfranchise individuals, and that we do everything we can to rehabilitate and reintegrate offenders back into the community. In Chapter 6 of the book, I argued that respect is one of the six dimensions of human well-being that it is the job of justice to secure a sufficient level of, and that it is completely compatible with free will skepticism. Unlike retributivists, I do not think respect for persons or concern for their well-being ends at the point of wrongdoing. Retributivists, however, appear to have a limited technical understanding of respect for human dignity in mind, one which focuses on giving wrongdoers their just deserts. They see this as the proper way to respect individuals as morally responsible agents. My account, on the other hand, is informed by the capabilities approach to justice and the public health approach to criminal behaviour, both of which view respect more holistically. Respect, as I conceive it, is an essential element of human flourishing. Well-being is therefore set back whenever individuals are perceived as being of lesser value because of membership in a particular group or because they are judged by their worst act and not their humanity. This is easy to see when invidious judgments are made about people based on race, gender, or economic class. But I also contend that the retributivist tendency to view offenders as simply 'criminals' or 'felons', and therefore deserving of punishment, stigmatization and disenfranchisement, is also a form of disrespect and harmful to human flourishing.

While retributivists like to see themselves as the champions of human dignity, in actual fact the practices they permit, and the violations of human dignity they encourage, are often paradigmatic examples of disrespect. Individuals are wronged and their well-being diminished when they are subject to punitive practices that dehumanize, stigmatize and disenfranchise them. On any neutral understanding of dignity and respect, such punitive practices would be considered textbook violations of human dignity.

On my model, all individuals are worthy of respect, including those we incapacitate. My model provides no justification for dehumanizing, disenfranchising or treating cruelly the individuals we must incapacitate. In fact, my model requires that we consider the well-being of those individuals and do everything we can to rehabilitate and reintegrate them back into society. The public health-quarantine model also endorses Norway's *principle of normality*,[72] which maintains that:

- Incapacitation is the restriction of liberty, but no other rights have been removed by the sentencing court. Therefore, the sentenced offender has all the same rights as all other citizens.
- No-one shall serve their sentence under stricter circumstances than necessary for the security in the community. Therefore, offenders shall be placed in the lowest possible security regime.
- During the serving of a sentence, life inside will resemble life outside as much as possible.

I view the principle of normality as a form of respect for human dignity, one that the retributivist fails to acknowledge or extend to criminal wrongdoers. Instead, the

72. For an official statement of the principle, see: <www.kriminalomsorgen.no/information-in-english.265199.no.html>.

retributivist conception of human dignity is like a parent who, while spanking their child, says: 'I'm not doing this for my own sake, I'm doing it to respect you as a morally responsible agent by giving you your just desert'. Hogwash! This is a perverse notion of human dignity and should have never gained traction in the first place.

Lastly, I contend that free will skeptics can additionally justify the safeguards needed to protect the rights of offenders and accused people in virtue of the fact that they are persons. In Chapter 6, I argue that there is no inconsistency in accepting a Kantian regard for respect, dignity, and treating individuals as ends-in-themselves, without accepting Kant's particular attitudes on free will.[73] This is because the capabilities account of social justice maintains that respect is something of independent moral significance regardless of whether individuals possess free will, since it matters centrally to everyone, whatever the particular life plans and aims each has. Appealing to Darwall's[74] distinction between 'appraisal respect' and 'recognition respect', and Vilhauer's distinction between 'action-based' desert claims and 'personhood-based' desert claims,[75] I argued that free will skeptics are able to preserve the latter of each pair even if they must reject the former. This is because action-based desert claims may presuppose the existence of free will, while personhood-based desert claims arguably do not. 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'.[76] I also argued that respect for persons simply follows from the fact that it is an essential dimension of well-being. As Nussbaum states, 'Governments must treat all people respectfully and should refuse to humiliate them'.[77] This is because of the 'centrality of notions of dignity and respect in generating the entire capabilities list'.[78] On the capabilities approach, respect and dignity are centrally important for establishing political principles that can provide the grounding for constitutional law and public policy in a nation aspiring to social justice. Hence, the public health-quarantine model maintains that 'human dignity, from the start, is equal to all who are agents in the first place.'[79]

The public health-quarantine model need not limit itself, then, to the safeguards provided by the principle of least infringement, the conflict resolution principle, and the non-retributive principle of proportionality discussed earlier. It can also appeal to a robust notion of respect for persons in arguing for various safeguards since this is perfectly consistent with the rejection of basic desert moral responsibility. There is no reason to think, then, that giving up on basic desert moral responsibility will lead to

73. See Pereboom, *Living Without Free Will* (n 69) 150–52; B Vilhauer, 'Free Will Skepticism and Personhood as a Desert Base' (2009) 39 Canadian Journal of Philosophy 489; B Vilhauer, 'The People Problem', in GD Caruso (ed), *Exploring the Illusion of Free Will and Moral Responsibility* (Lexington Book 2013); B Vilhauer, 'Persons, Punishment, and Free Will Skepticism' (2013) 162 Philosophical Studies 143; E Shaw, 'The Implications of Free Will Skepticism for Establishing Criminal Liability', in E Shaw, D Pereboom and GD Caruso (eds), *Free Will Skepticism in Law and Society: Challenging Retributive Justice* (Cambridge University Press 2019).
74. S Darwall, 'Internalism and Agency' (1992) 6 Philosophical Perspectives 155.
75. See references in n 72.
76. Vilhauer, 'The People Problem' (n 73) 148.
77. M Nussbaum, *Creating Capabilities: The Human Development Approach* (Harvard University Press 2011) 26.
78. ibid.
79. ibid 31.

abuses of human dignity, disproportionate punishment, or the absence of due process safeguards. The public health-quarantine model has the resources needed to protect against these concerns.

One final point. Sifferd also raises concerns about rehabilitation and autonomy – and provides the forced chemical castration of sex offenders as a potential problem-case for my account. In reply, I would direct readers to Section 8.3 of the book, where I address concerns about reasons-responsiveness, rehabilitation and autonomy at length. In particular, I argue *against* the forced chemical castration of sex offenders – but do leave open the option on a *voluntary basis*, 'especially if it is carried out in a way that respects autonomy by leaving the decision up to the individual'.[80] I also propose, along with Pereboom, 'that rehabilitative methods that directly appeal to a criminal's rational capacities should be preferred and attempted first'.[81] When these fail, we contend that

> it is sometimes acceptable to employ therapies that mechanically increase an agent's capacities for reasons-responsiveness, but that these therapies should involve the participation of the subjects to the greatest extent possible (e.g., talking therapies in conjunction with other forms of treatment), should involve the consent of the subject, and should be ordered such that noninvasive methods are prioritized.[82]

A common example of the latter includes the use of medication to counteract addiction or depression. It would seem mistaken to claim that such a mode of treatment is illegitimate because it circumvents capacities for rational and autonomous response. In fact, this sort of treatment often produces responsiveness to reasons where it was previously absent.[83] The approach to rehabilitation I advocate therefore has more in common with Sifferd's than she acknowledges, since both stress the need for consent, respect for autonomy, and the importance of restoring and appealing to an agent's reasons-responsive capacities.

4 ALEC WALEN

Alec Walen is a well-known retributivist and the author of the *Stanford Encyclopedia of Philosophy* entry on 'Retributive Justice'. In his comments here, however, he chooses not to defend the retributive account of punitive hard treatment but instead focuses his attention on my *Skeptical Argument* against retributivism. In fact, he's the only one of the six participants to seriously and directly address my skeptical arguments. In particular, he (a) sets out to defend compatibilism against the manipulation argument. He also objects that my argument against retributivism (b) 'proves too much' and that (c) I am operating with a 'misguided conception of the basis for blame'. As with the other participants, Walen's objections are both challenging and insightful. I thank him for engaging so thoroughly with my view and will try my best to respond to each of his objections below, beginning with his defence of compatibilism.

80. Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (n 1) 291.
81. ibid. See also Pereboom and Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life' (n 5).
82. Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (n 1) 291.
83. See Pereboom, *Living Without Free Will* (n 69).

4.1 Walen's Compatibilist Response

Walen acknowledges that 'Caruso's objections to libertarianism are thoroughly convincing',[84] and instead sets out to defend a version of compatibilism according to which 'we are free in the relevant sense if we can control our actions in the right way'. He correctly notes that I am a *source incompatibilist* and that my core argument in support of this view is the manipulation argument (although he overlooks my other main argument against compatibilism – the luck pincer). He begins by summarizing the argument using Pereboom's famous four-case version, which sets out three examples of actions that involve manipulation.[85] The first features the most radical sort of manipulation consistent with the proposed compatibilist conditions and with intuitive conditions on agency, each progressively more like the fourth, which is a case of ordinary determinism where the agent's action is causally determined in a natural way. The argument maintains that just as manipulated agents are not morally responsible in the basic desert sense, since their actions are ultimately the result of factors beyond their control, neither are agents in the ordinary case of determinism. All four cases involve Professor Plum who decides to murder White for the sake of some personal advantage and succeeds in doing so. The challenge for the compatibilist is to point out a relevant and principled difference between any two adjacent cases that would show why Plum might be morally responsible in the latter example but not in the earlier one. Proponents of the manipulation argument maintain that this cannot be done.

Walen's replies to this argument by offering what is known as a *hard-line* reply,[86] which maintains that 'it is not so clear that Plum should not be held responsible for the murder even in Case 1'. In support of this claim, he introduces Case 1*, 'in which Plum is wavering and unready to kill White, and then Jones encourages him to kill White, telling him that he'd be a coward if he didn't'. Walen maintains that, '[t]he law would take that to be no excuse', and adds that 'the problem for Plum is that he was willing to murder with such a small nudge'. He goes on to claim that, '[t]he same holds, however, in Case 1, and that invites the thought that he is equally culpable in Case 1'. In the end, he concludes that 'Plum is culpable in Case 1 because when the motive to murder was triggered in him he did not have a sufficiently robust commitment to avoid acting on it'. He acknowledges that 'it is tempting to excuse him in this case', but thinks 'the temptation is not compelling'.

While this is an interesting and novel way of responding to the manipulation argument, I do not think it succeeds. First, I would say that hard-line approaches in general conflict too deeply with our intuitions about *sourcehood* and the relevant class of manipulations. Many people find it highly implausible that Plum, say in Case 1 – where a team of neuroscientists directly manipulate Plum by producing in him a neural state that realizes a strongly egoistic reasoning process that deterministically results in his decision to kill White – could be morally responsible in the basic desert sense for his action given how it came about. And it is not just a matter of intuition that leads us to conclude this.

84. A Walen, 'Determinism, Compatibilism, and Basic Desert' (2021) Journal of Legal Philosophy (this issue), n 4.
85. See Pereboom, *Living Without Free Will* (n 69); and Pereboom, *Free Will, Agency, and Meaning in Life* (n 40).
86. See M McKenna, 'A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument' (2008) 77 Philosophy and Phenomenological Research 142.

The incompatibilist can argue, as Neal Tognazzini suggests, that ‘far from presupposing an incompatibilist conception of sourcehood, the manipulation argument is meant to give us reason to adopt that conception’.[87] According to Tognazzini, the dialectic of such an argument runs as follows:

> The incompatibilist presents an allegedly disturbing manipulation scenario [or a series of scenarios], and then says: ‘Whatever sourcehood is at the end of the day, this guy clearly doesn’t have it, and since sourcehood is required for moral responsibility, this guy ain’t morally responsible either.’ It’s this thought which gets regimented as [the first premise in manipulation arguments], the claim that the manipulated agent is not morally responsible for acting on his implanted psychological states.

Then the incompatibilist points out that there seems to be no difference between the manipulated scenario and a plain old deterministic world, at least as far as responsibility is concerned.

The incompatibilist then concludes that since the agent in the manipulated scenario isn’t responsible for his action, and since a plain old deterministic world isn’t relevantly different, the merely determined agent isn’t responsible for his action either. And since the problem with the manipulated agent was a lack of sourcehood, that must also be the problem with the merely determined agent. Thus, the proper understanding of sourcehood must be one that is incompatible with determinism.[88]

I agree with Tognazzini that the judgment of non-responsibility in the first premise is not inferred from the question-begging premise that sourcehood is incompatible with determinism; rather it is inferred from the intuition that, whatever sourcehood requires, it cannot be had by the agent in the manipulated scenario (together with the claim that sourcehood, whatever it requires, is itself required for moral responsibility). In other words,

> the proponent of the manipulation argument wants to stay neutral, at first, on the correct conception of sourcehood, and the manipulation scenario is supposed to serve up a desideratum of any adequate conception of sourcehood: ‘I think we’d all agree’, the incompatibilist says, ‘that this guy isn’t the source of his action’.[89]

Second, the hard-line approach generally adopts an *initial attitude* toward Case 4, the case of natural determinism, and then proceeds to argue in the opposite direction that manipulation cases are no threat to free will and moral responsibility. There are, however, a number of different initial attitudes one can bring to the ordinary deterministic example. As Pereboom identifies them, the *resolute compatibilist* response is to ‘deny that under these circumstances causal determination poses even a prima facie threat to our everyday assumption, and that it is rational to refuse to take seriously any further consideration for there being such a threat’.[90] A distinct approach affirms that causal determination provides a reason for giving up the responsibility assumption but claims that so far the issue has not been settled. This is the *neutral inquiring* response. ‘By this response it is initially epistemically rational not to believe that the agent in an ordinary deterministic example is morally responsible in the basic desert sense, and not to believe that he isn’t, but to be open to clarifying considerations that would

87. Neil Tognazzini, ‘The Structure of a Manipulation Argument’ (2014) 124 Ethics 358.
88. ibid.
89. ibid.
90. Pereboom, *Free Will, Agency, and Meaning in Life* (n 40) 93.

make one or other of these beliefs rational'.[91] It is crucial to note that the neutral inquiring attitude differs significantly from that of the *confirmed agnostic*. The confirmed agnostic claims that it is not clear that the ordinary causally determined agent is morally responsible in the sense at issue, and that it is not clear that he is not, but, like the resolute compatibilist, maintains that it is rational to consider enquiry into the issue closed, 'and for this reason it is not open to further clarifying considerations'.[92]

I agree with Pereboom that the confirmed agnostic response, the response that generates the hard-line conclusion, is not the appropriate response to take. Instead, the most attractive way of conceiving manipulation arguments involves supposing that the neutral inquiring attitude about ordinary determined agents is initially epistemically rational. The reason the neutral inquiring attitude is the appropriate attitude to adopt is that it allows for clarifying considerations to alter our thinking and 'it's the best one for the opposing parties in the debate to make if there is to be a productive engagement'.[93] But once we adopt the neutral inquiring attitude, we can see how an analogous manipulation case functions as a clarifying consideration that makes rational the belief that the ordinary causally determined agent is not morally responsible.

Now, Walen could argue that he does, in fact, adopt the neutral inquiring attitude but that his Case 1* provides its own 'clarifying considerations' and countervailing reasons for concluding that Plum should be considered morally responsible even in Pereboom's Case 1, where the team of neuroscientists directly manipulate Plum's neural states. But I see at least two problems with this reply. First, Case 1* is nothing like the direct manipulation involved in Case 1 of Pereboom's four-case manipulation argument. As a result, whatever intuitions it generates would not carry over. Instead, Case 1* is more akin to Pereboom's Case 4, the case of ordinary determinism. The only real difference is that Walen has spelled out one relevant antecedent determinant of Plum's decision, the encouragement and taunting of Jones. On the assumption of determinism, however, *all* of our decisions and actions would be causally determined by antecedent event, whether those events include the nudging of a friend, the advice of a doctor, or the circumstances and experiences of one's past.

Let's say, for instance, I'm addicted to chocolate, and so whenever there's chocolate in the house, I end up eating it all. My desire for chocolate, we can say, is due to various neural, psychological and social determinants. One day, however, I decide to join Chocoholics Anonymous and after attending a bunch of Chocoholics' meetings I eventually gain the ability to resist my desire for chocolate. Since compatibilists retain the assumption of determinism in order to reconcile it with free will, they would have to admit that *the desire to stop desiring chocolate*, which resulted in my decision to join Chocoholics Anonymous, was *itself* causally determined by factors beyond my control – e.g., the cumulative effects of advice from my doctor, comments made by my wife about my weight, things I read in the newspaper, and the like. I see Jones's nudging of Plum in Case 1* as akin to my doctor telling me that I need to give up chocolate or my wife encouraging me to give up sweets because I'm gaining weight – each of these antecedent determinants are part of a causal chain that, under the assumption of determinism, result in whatever action I end up doing. Walen's example therefore begs the question against the incompatibilist by simply assuming that ordinary determinism is no threat to basic desert moral responsibility. In fact,

91. ibid.
92. ibid.
93. ibid.

Walen's position seems more akin to the *resolute compatibilist* that adopts an unapologetic hard-line approach than someone who adopts the attitude of a neutral inquiring. Since Case 1* is actually analogous to Pereboom's Case 4, it does not help to reason in the opposite direction and argue that *since Plum is morally responsible in the case of natural determinism*, Plum's responsibility must transfer to Case 1 since there is no relevant difference between agents in Cases 4 and 3, 3 and 2, and 2 and 1. Such reasoning is clearly question begging.

Second, Walen's defence of compatibilism fails to consider the luck pincer and how the pervasiveness of luck undermines free will and basic desert moral responsibility. I mention it here, not to knock Walen for failing to address it in the extremely limited space he had, but because it further reveals a weakness in his response to the manipulation argument. When Walen points to the fact that Plum, in Case 1*, was 'willing to murder with such a small nudge' as the relevant feature that makes him morally responsible – and similarly that, 'Plum is culpable in Case 1 [of Pereboom's four-case manipulation argument] because when the motive to murder was triggered in him he did not have a sufficiently robust commitment to avoid acting on it' – I would argue that he's overlooking the importance of constitutive and present luck. This is because one's susceptibility to influence and manipulation is itself the by-product of constitutive luck, present luck, or both. For Walen's response to the manipulation argument to succeed, then, he would need to overcome not only my first objection but also the luck pincer.

4.2 Proving Too Much

Walen's second objection is that my skeptical argument proves too much. In particular, he argues that I cannot preserve the forward-looking *answerability* account of responsibility I endorse, nor can I hang on to the idea that it is hard to justify 'using people merely as a means'. The problem, he maintains, with a purely forward-looking account of responsibility grounded in future protection, future reconciliation, and future moral formation is that, 'if the calling [to account of wrongdoers] isn't tied to what is fair given what the agent has done, then there's no inherent constraint on doing it only for wrongdoers'. Furthermore, he maintains that on my account, 'one can never be bound by a contract'. Let me address each of these concerns in turn.

With regard to the irrelevance of wrongdoing, I would first point out that free will skeptics need not reject *axiological judgments* of right and wrong, good and bad, behaviour. Even if we came to hold that a serial killer was not blameworthy due, let's say, to a degenerative brain disease, skeptics contend that we could still justifiably agree that his actions are morally bad. Judgments of moral goodness and badness need not require an agent who is blameworthy or praiseworthy, they simply require grounds by which we can differentiate between the two types of judgments. If one were a Calvinist, for example, they could point the transcendent moral law as a way to judge while simultaneously rejecting all moral responsibility.[94] Less exalted moral systems, such as utilitarianism or Kantianism, provide alternative ways of grounding moral judgments. Of course, if one were to adopt a Kantian test of universalizability while rejecting the rest of Kant's moral views (which do presuppose

94. See Bruce N Waller, 'Virtue Unrewarded: Morality without Moral Responsibility' (2004) 31 Philosophia 3.

agents are morally responsible), it would hardly be an orthodox Kantian view. But, as several skeptics have noted, the denial of moral responsibility is not inconsistent with the principles of Kantian moral rationalism.[95] It is arguable, then, that axiological judgments of moral goodness and badness would not be affected by moral responsibility skepticism, and this may be sufficient for moral practice.

Second, I would point out that, while Pereboom's answerability/conversational account of moral responsibility is, in fact, forward-looking, it does retain at least one important, but uncontroversial, backward-looking feature – i.e., that agents must be *causally responsible* for doing some bad or wrongful act before they can be appropriate targets of moral protest and exchange. 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 moral exchange licensed by the forward-looking account of moral responsibility would only seem appropriate when some agent is causally responsible for some wrongdoing, since it is only then that calling her to account would benefit future protection, reconciliation and/or moral formation. If Jack, for instance, is the one who steals his neighbour's newspaper every morning, it would be wrong on this account to invite Jill to 'evaluate critically what her actions indicate about her intentions and character, to demand an apology, or request reform'[96] since doing so would not aid future safety, reconciliation or moral formation – the reason being that Jill was *in no way* causally involved in the wrongdoing. Hence, Walen is mistaken that on the forward-looking account of moral responsibility there is 'no inherent constraint' on only calling wrongdoers to account.

With regard to contracts, Walen provides the following example:

> Suppose A signs a contract with B to deliver 100 widgets, and then A discovers that it is hard to get widgets and wants out of the contract, so she breaches. B sues for breach. Now imagine A brings a Case 1 defense: 'I signed the contract, it's true, but I was manipulated by neuroscientists to do so'. Just as a court would be tempted to give Plum a Case-1 defense to a murder charge, so too a court would be tempted to give A 'lack of capacity' or 'undue influence' defense in Case 1 contract cases. But if Case 1 and normal life are on par, as Caruso thinks, then one can never be bound by a contract.[97]

I find this example odd for two reasons. The first has to do with nature of contract law and how it differs from criminal law, and the second has to do with the specifics of the example.

First, contract law, much like tort law, could demand restitution for breach of contract or provide for the damages to the injured party *without* invoking basic desert moral responsibility. This is because retributive punishment, in the context of criminal law, requires a notion of basic desert, whereas contract law and tort law need not. Some wrongful conduct can be a tort but not a crime, and vice versa. A *tort* is something that occurs when one person's negligence directly causes property or personal damage to another individual – and it need not be intentional. An example of an unintentional tort would be someone being injured by a faulty product or someone's pet. A crime, on the other hand, is generally an intentional, planned wrongdoing. As general rule, in tort

95. See Vilhauer, 'The People Problem' (n 73); Vilhauer, 'Persons, Punishment, and Free Will Skepticism' (n 73); Pereboom, *Free Will, Agency, and Meaning in Life* (n 40).
96. Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (n 1) 38.
97. Walen, 'Determinism, Compatibilism, and Basic Desert' (n 84).

law, the financial harm suffered by the victim as a result of a tort is the only issue. The same is generally the case in contract law, which governs making contracts, carrying them out, and fashioning a fair remedy when there's a breach. Tort and contract law both attempt to adjust for harms done by awarding damages to a successful plaintiff who demonstrates that the defendant was the cause of the plaintiff's losses. Criminal law, on the other hand, is concerned with more than the mere restitution and compensation of victims. It is also concerned with punishing wrongdoers for their criminal acts, not just as an act of restitution but as an expression of the state's disapproval of both the offense and the offender. As a free will skeptic, I maintain that contract law and tort law are consistent with the rejection of basic desert moral responsibility, since the restitution and compensation of victims can be justified by appealing to the rights of those harmed and notions of responsibility available to the skeptic. That is, contract law and tort law need not assume agents are blameworthy and morally responsible in the basic desert sense, only that they are causally responsible for some breach of contract or negligent act that caused harm and are therefore responsible, in the *civil liability* sense, for compensating the victim(s).

Second, I think Walen's example is faulty because the conditions necessary for entering a valid contract are different from the conditions necessary for free will and basic desert moral responsibility. Note that if Plum's defence in Case 1 succeeds in absolving him of criminal guilt, *and* the manipulation argument is correct, then in the case of ordinary determinism *no* one is truly deserving of punishment since no one is morally responsible in the required sense. The same is not the case with contracts. A's defence in Case 1, even if successful, would *not* entail that 'one can never be bound by a contract' – it would only entail that those individuals manipulated into entering a contract by a team of neuroscientists would not be bound to that contract. Determinism does not threaten the binding power of a contract – nor for that matter does indeterminism or luck – since contracts do not presuppose the existence of free will or basic desert moral responsibility in the same way that criminal law does.[98] Legally binding contracts require only that capable parties mutually, consensually and voluntarily assent to the terms of the agreement, not that they have the control in action required for basic desert moral responsibility. I'll leave it to others, then, to decide whether external manipulation by neuroscientists in cases where the agent

98. The criminal law is clearly founded on the idea that persons can be held morally responsible for their actions because they have freely chosen them. The US Supreme Court, for instance, has asserted: 'A "universal and persistent" foundation stone in our system of law, and particularly in our approach to punishment, sentencing, and incarceration, is the "belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil"' (*United States v Grayson*, 438 U.S. 41 at 52 (1978), quoting *Morissette v United States*, 342 U.S. 246, 250 (1952)). Indeed, the US courts have observed that '[t]he whole presupposition of the criminal law is that most people, most of time, have free will within broad limits' (*Smith v Amontrout*, 865, F.2d 1502, 1506 (8th Cir. 1988), and that 'the law has been guided by a robust common sense which assumes the freedom of will as a working hypothesis in the solution of its problems' (*Steward Machine Co. v Davis*, 301 U.S. 548, 590 (1937)). The US Supreme Court, in fact, has gone so far as to suggest that 'a deterministic view of human conduct … is inconsistent with the underlying precepts of our criminal justice system' (*United States v Grayson*, 438 U.S. 41 at 52 (1979)). While this last claim in controversial, since some legal scholars claim that the criminal law only requires compatibilist free will, one thing is clear: If human beings lack the control in action, that is, the free will, required for basic desert moral responsibility, then our current conception of the criminal law will need to be revised.

approves of their own motivational states, is reasons-responsive, etc., undermines the consent needed for contracts.[99] Whether or not it does, I maintain that there are relevant differences between the conditions necessary for free will and those for entering valid contracts such that the former, but not the latter, is inconsistent with ordinary determinism. If this is correct, then under normal conditions individuals can be, and usually are, bound by the contracts they mutually and consensually enter into.[100]

4.3 Conceptions of Blame

Turning now to Walen's final objection, he takes issue with my claim that retributivism presupposes that wrongdoers deserve punishment because they 'have knowingly done wrong'. He claims that this view naturally raises the question: Why would anyone knowingly do wrong? According to Walen, '[t]his is a problem that goes back to Plato's Protagoras, and I think it has no good answer'. Instead, he argues that 'the better account of blameworthiness bases it on doing wrong because one lacked sufficient care for the rights of others'. This analysis, he claims, turns on a very particular counterfactual: 'if one had sufficient care for the rights of others, would one have acted appropriately given what one knew and given one's ability to resist psychological pressure to do otherwise'. If so, he claims 'one's failure is a culpable manifestation of insufficient moral concern for others'. Walen's account is therefore a version of a quality of will conception of blameworthiness.

In general, I have no issue with Walen's account of blameworthiness. Grounding blameworthiness in insufficient moral concern is a plausible and common move for compatibilists. The core philosophical question, however, remains: Are we justified in blaming someone for expressing insufficient moral concern if that lack of concern was itself the result of factors beyond the control of the agent (e.g., determinism, indeterminism, or luck)? The skeptic says 'no' (because of either the manipulation argument, luck pincer, consequence argument, no-forking-paths argument, or some combination), while the compatibilist says 'yes'. Providing an account of blameworthiness grounded in concern for the rights of others, does not *itself* answer that question. For a compatibilist account of blameworthiness to succeed, it would need to overcome the various incompatibilist arguments – and I have argued that Walen has not successfully defeated the manipulation argument, let alone the luck pincer, which he fails to address. I therefore wholeheartedly disagree with Walen that all we 'need for basic desert of censure or blame' is the 'ability to appreciate that others have claims on one, that one has reasons to respect those claims, and that if one fails to do so, one is not the sort of person one should want to be and should feel guilty for what one has done'.

99. One could, for instance, defend a hard-line reply in case of contracts, while still rejecting the hard-line reply in the case of free will, by pointing to the different requirements of each.
100. Walen also claims that my skeptical argument proves too much with regard to my prohibition on manipulative use, writing: 'Caruso is even mistaken that he can hang onto the idea that it is hard to justify "using people merely as a means"' (Walen, 'Determinism, Compatibilism, and Basic Desert' (n 84)). He claims: 'The problem is that that notion is tied to the thought that people can waive or forfeit their rights not to be used as a means. Get rid of the relevance of past acts, and that sort of qualification of the claim not be used as a means is gone. But without it, the restriction is implausibly strong' (ibid). It's simply wrong, however, to claim that my view dismisses the relevance of past acts. I see no reason for thinking that adopting the skeptical perspective entails that prior consent becomes irrelevant. And since Walen himself fails to provide me with such a reason, I will simply set this objection aside.

5 DERK PEREBOOM

Derk Pereboom has been my inspiration and partner in developing the public health-quarantine model[101] and has influenced my arguments for free will skepticism more than any other philosopher. Generally, there is very little daylight between our positions, so when Pereboom disagrees with me, I tend to get concerned. Recently, Pereboom has departed slightly from the core of the public health-quarantine model in response to criticisms from general deterrence theorists.[102] In this section, I will explain why I do not follow Pereboom in that departure and instead prefer to keep the model free of any justificatory appeals to general deterrence. Largely in response to considerations raised by Victor Tadros,[103] who has argued that concerns about manipulative use can sometimes be overridden,[104] Pereboom has conceded that perhaps a greater level of general deterrence might be desired than what is permitted by the right of self-defence. In particular, he has argued that with regard to 'monetary fines and short prison terms'[105] it might sometimes be justified to 'use unfree wrongdoers in ways that involve such penalties to subserve general deterrence'.[106]

5.1 Specific and General Deterrence

Before explaining why I disagree with Pereboom's recent acceptance of limited punitive measures, we need to take a step back and recall the distinction between *special deterrence* and *general deterrence*. General deterrence is the deterrence achieved from the threat of legal punishment on the public at large. Special deterrence, on the other hand, is aimed at previous offenders in order to reduce the likelihood of their re-offending. Pereboom argues that it is significantly easier to ground a limited form of special deterrence in the right to harm in self-defence and defence of others than is general deterrence. This is because the use of incapacitation qualifies as a case

101. Pereboom, *Living Without Free Will* (n 69); Pereboom, *Free Will, Agency, and Meaning in Life* (n 40); Pereboom and Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life' (n 5); G Caruso and D Pereboom, 'A Non-Punitive Alternative to Punishment', in F Focquaert, B Waller and E Shaw (eds), *Routledge Handbook of the Philosophy and Science of Punishment* (Routledge 2020).
102. D Pereboom, 'A Defense of Free Will Skepticism: Replies to Commentaries by Victor Tadros, Saul Smilansky, Michael McKenna, and Alfred R. Mele' (2017) 11 Criminal Law and Philosophy 617; D Pereboom, 'Free Will Skepticism and Preventions of Crime', in E Shaw, D Pereboom and GD Caruso (eds), *Free Will Skepticism in Law and Society: Challenging Retributive Justice* (Cambridge University Press 2019); D Pereboom, 'Incapacitation, Reintegration, and Limited General Deterrence' (2020) 31 Neuroethics 87.
103. V Tadros, *The End of Harm* (n 24); and V Tadros, 'Doing without Desert' (2017) 11 Criminal Law and Philosophy 605.
104. Tadros has argued that while there are a wide range of cases where it is intuitively objectionable to use someone manipulatively, there are also exceptional cases where the prohibition on manipulative use can be overweighed. Tadros rejects basic desert but nevertheless aims to justify certain forms of punishment that are more severe than those that are justified on the grounds of self-defense. To justify these additional measures, he develops a theory that claims that the manipulative use objection can be answered by invoking duties that wrongdoers owe to victims (see Tadros, *The End of Harm: The Moral Foundations of Criminal Law* (n 24)).
105. Pereboom, 'Free Will Skepticism and Preventions of Crime' (n 102) 104.
106. Pereboom, 'Incapacitation, Reintegration, and Limited General Deterrence' (n 102) 7.

of threat-elimination (or eliminative harming) and not as manipulative use.[107] Pereboom argues that the use of preventive detention is justified 'as special deterrence on the ground of the right to self-defence',[108] as long as it is in accordance with the principle of least infringement and the minimum harm required to protect ourselves and others. In this way, Pereboom justifies special deterrence, understood not as a form of punishment but as the right to use preventive detention and eliminative harming, by appealing to the right of self-defence and defence of others.

But what about general deterrence? Even if incapacitation and eliminative harming can be justified as special deterrence on the ground of the right to self-defence, does not general deterrence violate the prohibition on manipulative use? Here, Pereboom provides two separate answers – the first that I accept, and think is sufficient, and the second that I reject. Pereboom first correctly notes that a significant level of general deterrence will result as a natural side effect of a system based on incapacitation. He calls this 'free general deterrence'. Free general deterrence follows from the state's requirement to be transparent about its practices regarding the incapacitation of the seriously dangerous. Importantly, though, it comes with 'a significant limitation on how much harm can legitimately be inflicted – only the minimum harm required to protect against an aggressor can be justified'.[109] Since it does not extend beyond the minimum harm required to eliminate a threat on the right of self-defence, it does not violate the prohibition on manipulative use.

While I agree with Pereboom that incapacitation justified on the right of self-defence will naturally produce free general deterrence, I contend that this is *all* we should seek to justify. I therefore disagree with Pereboom that we should seek to go beyond free general deterrence and justify 'more exacting general-deterrence-subserving penalties'.[110] But before explaining why I disagree, we should first get clear on Pereboom's position. In a series of recent papers,[111] Pereboom has proposed a way to justify a 'limited degree' of general deterrence that surpasses free general deterrence. As he describes it:

> First, I've argued that it's plausibly the right to life, liberty, and physical security of the person that have a key role in making the use objection to general deterrence intuitive. Those rights are grounded in the more general right to a life in which one's capacity for flourishing is not compromised in the long term. There is a heavily weighted presumption against punishment as use where that involves intentional killing, long-term confinement, and infliction of severe physical or psychological harm. But if the proposed penalties are significantly less extreme, such as monetary fines and short prison terms, it would be permissible to use unfree wrongdoers in ways that involve such penalties to subserve general deterrence.[112]

According to Pereboom, there 'may well be many circumstances in which effective general deterrence would require modest penalties of these sorts' – that is, penalties that 'involve the imposition of more harm than can be justified on special deterrence grounds; that is, as free general deterrence'.[113] In such circumstances, Pereboom argues that it would be permissible to increase the severity of the penalty since

107. ibid 4.
108. Pereboom, 'Free Will Skepticism and Preventions of Crime' (n 102) 103.
109. ibid.
110. ibid.
111. See n 103.
112. Pereboom, 'Free Will Skepticism and Preventions of Crime' (n 102) 104.
113. ibid.

'monetary fines … don't hinder the prospects for a life lived at [a] reasonable level of flourishing'.[114] On Pereboom's proposal, general deterrence not justified on special deterrence grounds is 'plausibly justified if it substantially increases general deterrence value, and/or it substantially lowers the cost of deterrence, provided that the more exacting measure doesn't hinder the prospects for a life lived at a reasonable level of flourishing'.[115]

While I acknowledge that Pereboom's proposal is an attractive option for those free will skeptics who wish to justify more exacting measures that go beyond what is permitted by the public health-quarantine model, I resist it for three main reasons.

5.2 Initial Replies

My first reason for resisting Pereboom's proposal is that I think supplementing the public health-quarantine model with a limited form of general deterrence unnecessarily muddies the waters and sacrifices some of the distinct advantages the model has over rival accounts, including consequentialist deterrence theories. In my estimate, two of the most distinct features of the public health-quarantine model are that (a) it offers a nonpunitive approach to addressing criminal behaviour, not just a non-retributive one, and (b) it does not run afoul of the prohibition on manipulative use. But in one fell swoop, Pereboom's proposal sacrifices both of these features. By allowing in limited general deterrence as a justification for monetary fines and short prison terms, Pereboom has introduced a punitive component. As a result, his newly expanded theory must squarely face the *problem of punishment* presented by David Boonin, Michael Zimmerman, and others, who argue that *all* forms of punishment are unjustified.[116]

My version of the public health-quarantine model remains pure and free of any punitive components and is therefore compatible with the complete rejection of all justifications of punishment. I see this as an advantage, since the problem of punishment is a serious one and Boonin and Zimmerman may well be right that punishment is never justified. Pereboom's proposal also requires defending exceptions to the prohibition on manipulative use, even if only in limited cases, since general deterrence involves manipulative use. The version of the public health-quarantine model I defend, on the other hand, is consistent with an absolute prohibition of manipulative use, which, again, I see as an advantage.

Second, I think Pereboom underestimates the extent to which monetary fines and short prison sentences can, and often do, 'hinder the prospects for a life lived at a reasonable level of flourishing'.[117] Short prison sentences, for instance, can be extremely disruptive to the lives of those incarcerated as well as their family and friends. We know, for example, that prison sentences have a long-term and negative impact on a person's employment prospects[118] and even a short prison term will

114. ibid 105.
115. ibid.
116. See, e.g., Boonin, *The Problem of Punishment* (n 25); Zimmerman, *The Immorality of Punishment* (n 26).
117. For more on this point, see Section 9.2 of my book.
118. See, e.g., Ministry of Justice, 'Compendium of re-offending statistics and analysis' (2013) available at: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/278133/compendium-reoffending-stats-2013.pdf>.

cause an interruption in employment and increase the possibility of homelessness, bankruptcy, suicide, loss of healthcare, stigmatization and divorce.[119] Short prison terms also separate parents from children, which disrupts the development of positive relationships, and this can have long-term negative effects on children. Furthermore, short prison sentences are far less effective than community sentences at reducing re-offending.[120] In England and Wales, for instance, the reoffending rates for prison sentences of a year or less were at 65.5 percent in the last three months of 2016, rising to 67 percent for sentences of six months or less.[121] This is much higher than the 38 percent of people who re-offend after being served a court order, such as a community order or suspended sentence. As Christopher Kay correctly notes: 'Reoffending rates from short sentences have been a cause for concern for years.'[122] Meanwhile, 'research has shown the effectiveness of community-based sentences in reducing reoffending – at least compared to short prison sentences – and that community supervision is less likely to have a negative impact on employment and family time'.[123]

For these reasons, many experts now agree that we overuse short prison terms for nonviolent and persistent crime (Ministry of Justice 2018). David Gauke, the former Secretary of State for Justice in the United Kingdom, has, for instance, argued that short prison terms should only be used as a 'last resort'.[124] And Rory Stewart, the new Justice Minister, has called for a drastic reduction in the use of short prison sentences, writing:

> In March 2018 we launched our campaign that showed short sentences are short-sighted. We asked the government to review this issue and to consider introducing a new presumption against the use of short custodial sentences of less than six months. We also asked the government to strengthen community sentences so that they command public confidence and are better able to deal effectively with some of the underlying causes of persistent petty offending, including drug or alcohol addiction and mental ill-health.[125]

I endorse these recommendations and suggest that community sentences should be chosen over the use of short prison terms wherever possible. In fact, evidence from Scotland reveals that the presumption against prison terms of less than three months, introduced in 2010, has reduced use of short jail terms by 40 percent and in

119. Prison Reform Trust, 'Bromley Briefings Prison Factfile, (2018) available at: <http://www.prisonreformtrust.org.uk/Portals/0/Documents/Bromley%20Briefings/Autumn%202018%20Factfile.pdf>.
120. Ministry of Justice, 'Compendium of re-offending statistics and analysis'; Ministry of Justice, 'Offender management statistics quarterly: April to June 2018', (2018) available at: <www.gov.uk/government/statistics/offender-management-statistics-quarterly-april-to-june-2018>.
121. C Kay, 'Short prison sentences as a last resort won't work unless the probation service is fixed', (2019) The Conversation, available at: <https://theconversation.com/short-prison-sentences-as-a-last-resort-wont-work-unless-the-probation-service-is-fixed-110480>.
122. ibid.
123. ibid.
124. See, for instance: <www.thetimes.co.uk/article/under-a-year-in-jail-must-be-last-resort-says-justice-chief-david-gauke-msdbmfmbb>.
125. R Steward, 'Reducing the use of short prison sentences in favour of a smarter approach' (2019) Revolving Door Agency, available at: <www.revolving-doors.org.uk/file/2347/download?token=e9wtT41q>.

2016/2017 crime in Scotland had fallen more than 18 percent with a 26 percent fall in property crime.[126] Results like this have led the Revolving Door Agency to write:

> The evidence is clear. Short prison sentences are short-sighted because they disrupt family ties, housing, employment and treatment programmes, but they do not provide any meaningful rehabilitation. These sentences contribute to prison 'chum' and volatility. They are ineffective at tackling petty crime. We can do better and should adopt a smarter approach.[127]

While Pereboom may now be willing to consider the use of short prison terms to subserve general deterrence, I do not think he realizes the extent to which such penalties are ineffective and how they can, and often do, hinder the prospects for a life lived at a reasonable level of flourishing.

My third reason for rejecting Pereboom's proposal is that I think the public health-quarantine model already has the resources needed to deal with low-level, non-violent crime, and as a result there is no reason to sacrifice the principle of least infringement in an effort to accommodate limited general deterrence. We should be content with free general deterrence and only what is permitted by the right of self-defence and defence of others. It is not at all clear that seeking 'more exacting' penalties is either wise or justified. The right of self-defence and defence of others can, for instance, justify various liberty-limiting measures short of incapacitation and these can be used to address a host of nonviolent offenses like speeding, drunk driving and shoplifting. And this can be done without appeal to free will or basic desert. Driving privileges, for example, can be limited, restricted or removed for various reasons, including poor eyesight, age and medical conditions that make driving unsafe. Such restrictions are widely seen as justified on the grounds that they are necessary for the protection of society and the prevention of harm to others – and these measures are not forms of punishment nor are they implemented for the purposes of deterrence. By analogy, the public health-quarantine model can justify liberty-limiting policies for repeat speeders and drunk drivers since their actions manifest a blatant disregard for the safety of those around them. These restrictions can be implemented incrementally or all at once based on the nature of the violation and whether the individual is likely to re-offend.

In fairness to Pereboom, though, the replies I offered in section 9.2 of the book generally focused on specific crimes (like speeding and drunk driving) and cases where financial penalties and short prison sentences *would* 'hinder the prospects for a life lived at a reasonable level of flourishing'. In his comments here, he makes it clear that the kind of cases he has in mind, where the *general deterrence prerogative* would be justified, include 'manipulation of financial markets, large-scale embezzlement, and illegal use of political influence for gain in personal wealth and power'.[128] Those who commit such crimes, Pereboom writes:

> are typically not poor or from disadvantaged backgrounds, and so the public health model, as Caruso sets it out, is not calibrated to prevent crimes of this sort. Those who commit such crimes are often wealthy and well-educated, but willing to free-ride for reasons of self-interest. They are often good at calculating risk, at weighing the probability of detection against the probability of personal gain.

126. Scottish Government, 'Criminal Proceedings in Scotland 2017-18' (2019) available at: <www.gov.scot/publications/criminal-proceedings-scotland-2017-18/>.
127. See, for instance: <www.revolving-doors.org.uk/file/2347/download?token=e9wtT41q>.
128. D Pereboom, 'Non-free Deterrence' (2021) Journal of Legal Philosophy (this issue).

He goes on to write:

> The incapacitation invoked by free general deterrence would involve the threat of loss of political or professional standing, such as removal from political office or revocation of a license to trade in financial instruments. Whether threats of such incapacitation are sufficient to deter the crimes at issue is an empirical matter, but, in disagreement with Caruso (2021, Chapter 9), I would wager that they are often not. The penalties designed to address such crimes in place in the United States are typically already more severe than what free general deterrence would allow, and yet the incidence of these kinds of political and financial wrongdoing is fairly high. Reducing the strength of the deterrents is apt to increase the incidents of such wrongdoing. As noted, public health policy that aims to reduce poverty and improve access to health care and education is mismatched for crimes of this kind. These considerations motivate a turn to a non-free sort of general deterrence.[129]

In the following section, I will address these more specific concerns of Pereboom and argue that it remains a mistake to increase our punitive responses to white-collar crimes beyond what is permitted by the public health-quarantine model and the principle of least infringement.

5.3 White-Collar Crime

First, I agree with Pereboom that political and financial crimes differ in significant ways from violent crime. I acknowledge, for instance, that removal from public office or revocation of, say, a license to trade in financial instruments is often enough to protect society and guard against such offenders reoffending in the future. I also acknowledge that the preventative strategies needed to address such crimes require identifying additional social determinants of criminal behaviour. That said, I strongly caution against ramping up our punitive responses to political and financial crimes. After suffering through four years of the Trump administration, I understand the punitive impulse in such cases – whether it be motivated by a desire for retribution or general deterrence. Nevertheless, the commitments of my view, as well as my own reformist tendencies, recommend that, instead of *ramp up* our punitive responses to white-collar crimes to match those of violent crimes, we should *ramp down* our punitive responses to both. It's easy to recommend harsher treatment and 'more exacting' penalties for those who manipulate financial markets, embezzle money, commit large-scale tax fraud or illegally use their political influence, since our criminal justice system is already wildly unjust and excessively punitive for those less fortunate members of society, while white-collar criminals seem to be 'getting off easy'. Such thinking, however, only perpetuates injustice and increases our already excessively punitive responses to crime.

One of my main concerns with Pereboom's general deterrence prerogative is that once you allow in some 'limited degree' of general deterrence, it's hard to keep it 'limited' to only those offenders who are 'wealthy and well-educated' and (presumably) can afford some small fine or will not have their prospects for a life lived at a reasonable level of flourishing hindered by a short prison sentence. Several difficulties, I contend, confront Pereboom's proposal. For instance, I don't see how it would be just for the state to adopt a system of fines and short prison sentences that treated 'wealthy and well-educated' offenders more harshly than poor and less educated offenders since

129. ibid.

the former's ability to flourish would (presumably) not be hindered by such punishments.

Consider, for instance, the use of financial fines. Perhaps Pereboom could more easily argue that fines, unlike short prison terms, can increase general deterrence value without significantly hindering human well-being and flourishing. It is not clear, however, that this is true. To see why, consider the following example. A few years back I received a series of speeding tickets in a short period of time. It is not something I am proud of but I was doing a lot of long-distance travel at the time and I occasionally had a heavy foot on those long stretches of empty highway. Each ticket cost me roughly $350. The last violation put me in jeopardy of losing my license. The officer who pulled me over told me that I could contact the District Attorney and pay him $250 and he would petition the court to get the violation reduced or dismissed. So that is what I did. And as promised my speeding violation was reduced to a parking violation, saving me around a $150, points on my license, an increase in my insurance premiums, and potentially the loss of my driver's license.

Now, let us imagine the same scenario but with one small difference. In this case, let us assume the driver is someone who lives on the poverty line and who makes their living as an Uber or Lyft driver. Since they are unable to pay the District Attorney the $250, they obviously cannot pay the full $350 fine. They plead not-guilty and are assigned a court date but they are unable to make it since they cannot afford to take the day off from work. After not paying the fine, the court notifies them that their driver's license has been suspended. Unfortunately, the individual's only source of income comes from being a driver. They could stop driving but that will significantly hinder their 'prospects for a life lived at a reasonable level of flourishing'. So they continue driving with a suspended license. They're careful not to speed again but one day they are pulled over for a front headlight that has gone out. As a result, they are assessed even heavier fines. But being unable to pay the fines, both old and new, the court eventually issues a warrant for their arrest. From here we can easily imagine a series of cascading events that leads to the arrest of the individual – which, in turn, leads to additional lawyer fees, bail costs and possibly even jail time.

Here we have a conceivable, and dare I say common, set of circumstances where an apparently 'small' financial fine ends up hindering an individual's 'prospects for a life lived at a reasonable level of flourishing', since the cumulative costs incurred from the fine go well beyond the initial dollar amount and ultimately result in a significant diminishment in human well-being. As a result of the speeding tickets, this individual may lose their livelihood, their ability to pay their bills, and perhaps even their freedom. Small fines can have significant effects. As a recent article in *The Nation* explains: 'Debtor's prisons have been illegal in America since 1833. But that doesn't matter. We know about some ways people can languish in jail for being poor – if they cannot pay bail, for example, or if they rack up fines related to imprisonment that must be paid upon release'.[130] *Salon* magazine further explains:

> A symbol of Victorian England's inequitable nature made infamous by Charles Dickens, debtors' prisons were banned in the United States in 1833. The Supreme Court has affirmed the unconstitutionality of jailing those too poor to pay debts on three different occasions in the last century, finding that the 14th Amendment prohibits incarceration for non-payment

130. 'Prosecutors and judges have brought back debtors' prisons,' *The Nation*, 22 February 2018. Available at: <www.thenation.com/article/archive/prosecutors-and-judges-have-brought-back-debtors-prisons/>.

> of exorbitant court-imposed fines or fees without an assessment of a person's ability to pay and alternatives for those who cannot. 'Punishing a person for his poverty' is illegal, the Court said. Yet in recent years the modern-day equivalent of debtors' prisons have returned, as cities have grown to rely on punishing regimes of fines and fees imposed on their own residents as a major stream of revenue.[131]

The article goes on to write:

> Routine traffic tickets or even overdue student loan payments can set off a cycle of debt that also includes the suspension of a driver's license or professional license and, in some cases, jail time. A suspended driver's license makes it nearly impossible to get to work. When a person can't pay, courts add more fines on top of the original. If those fees aren't paid, a jail sentence is imposed.[132]

While I am certain Pereboom would oppose debtors' prisons and the inequalities they reinforce, I do not think he fully appreciates the extent to which fines and short prison sentences can, and often do, interfere with an individual's ability to achieve a reasonable level of flourishing. Financial fines compound other existing social injustices and can have disproportionate effects on the poor. I therefore maintain that we should resist Pereboom's proposal to supplement the public health-quarantine model with 'more exacting general-deterrence-subserving penalties', since these can exacerbate existing social inequalities and negatively impact human well-being.

At this point, Pereboom may fall back on his proviso that fines and short prison sentences *only be used* when it does not hinder the prospects for a life lived at a reasonable level of flourishing, like in those cases where an individual is 'wealthy and well-educated'. But, again, I do not see how it would be just for the state to adopt a two-tiered system of criminal justice, one that *prohibited* the use of general deterrence in the case of the poor and less educated, and one that *permitted* its use but only for the wealthy and well-educated. Furthermore, how would such a system be administered, practically speaking, and who would decide whether a particular fine or prison sentence sufficiently 'hindered the prospects for a life lived at a reasonable level of flourishing'? There's also the problem of determining *how much* wealth should be considered enough to justify general deterrence. Must an individual's income be over $100,000, $200,000, or is living above the poverty line sufficient? Lastly, how should we weigh the cost to well-being of spending a few months in jail against a net increase in general deterrence value? These are all serious concerns and I have yet to see how Pereboom's position resolves them.

I would also like to add that, on my theory, the responses available to political and financial wrongdoing are not limited to removing corrupt politicians from office and revoking one's trading license. As I argued in my previous set of replies, free will skepticism and the public health-quarantine model are consistent with tort law, which allows for the restitution and compensation of victims. That means that, without appealing to the justification of general deterrence, individuals like Bernie Madoff may still be required to compensate their victims for the money they stole. Such restitution would not be a matter of criminal law and general deterrence, but civil law and the rights of victims to be compensated for their losses. I therefore think Pereboom

131. 'A return of debtors' prisons: Jeff Sessions' war on the poor,' *Salon*, 29 December 2017. Available at: <www.salon.com/2017/12/29/a-return-to-debtors-prisons-jeff-sessions-war-on-the-poor/>.
132. ibid.

underestimates the tools already available to my model to address white-collar crime, as well as the ability of tort law to produce a significant amount of *free* general deterrence.

My final point has to do with the preventative aspect of my public health-quarantine model and how it can help obviate the need for general deterrence. Pereboom is correct that 'public health policy that aims to reduce poverty and improve access to health care and education is mismatched for crime of this kind [i.e., political and financial crime committed by wealthy and well-educated individuals]'.[133] That does *not* mean, however, that public health policies aimed at reducing the social determinants of white-collar crime would not prove effective. I acknowledge that in the book I focus on the social determinants of violence rather than, say, insider trading and embezzlement. That does not mean, however, that there are not common factors that make someone more susceptible to committing white-collar crimes. First off, it's important to note that the demographics of white-collar criminals are fairly consistent. Most tend to be white, married, middle-class, middle-aged men with some level of higher education and moderate to strong ties to their community, family and/or religion.[134] They also tend to commit their first white-collar crime in their late-thirties through mid-forties.[135] Second, white-collar crimes usually result from what is referred to as *theory of convenience* – a combination of motive, opportunity and willingness.[136]

There are several distinct factors that draw people to white-collar crime.[137] They include workplace acceptance of deviant behaviour; not having much to lose; getting high off risk; attractiveness to convenience (trying to make in five minutes what would normally take five years); a feeling of being wronged (people are more likely to harm others if they have been harmed themselves);[138] and narcissistic personality. Furthermore, convenience theory suggests that the extent of individual convenience orientation determines to what extent a person of respectability and high social status in the course of his/her occupation will make a decision to violate the law whenever alternative decisions are less convenient. For instance, a more stressful and greedy financial motive, and improved organizational opportunity to commit and conceal crime, and a stronger personal willingness for deviant behaviour have been shown to influence the extent of white-collar crime intention.[139]

Understanding the social determinants of white-collar crime provides us with the opportunity to adopt practices and policies that help obviate the need for general deterrence. These may include changing the culture of an institution, providing more workplace training and supervision, employing behavioural nudges that make it more convenient to follow the rules than to break them, screening for narcissistic

133. Pereboom, 'Non-free deterrence' (n 128).
134. See S Van Slyke, ML Benson and FT Cullen (eds), *The Oxford Handbook of White-Collar Crime* (Oxford University Press 2016).
135. ibid.
136. P Gottschalk, 'Convenience Triangle in White-Collar Crime: Case Studies of Relationships between Motive, Opportunity, and Willingness' (2018) 55 International Journal of Law, Crime and Justice 80.
137. See S Sarkis, 'Seven Factors that Draw People to White-Collar Crime' (2019) *Forbes*, available at: <www.forbes.com/sites/stephaniesarkis/2019/03/03/seven-factors-that-draw-people-to-white-collar-crime/?sh=3ed51d3c78f8>.
138. M Tsvetkova and MW Macy, 'The Social Contagion of Antisocial Behavior' (2015) 2 Social Science 36–49.
139. P Gottschalk, 'Theory of Convenience: Determinants of White-Collar Crime Intention' (2019) 41 Deviant Behavior 1431–39.

personality disorder, etc. I therefore fundamentally disagree with Pereboom when he claims, 'the public health model … is not calibrated to prevent crimes of this sort [i.e., white-collar crimes]'.[140] By focusing on the social determinates of violence and how a public health approach can identify, prioritize and address them, I may have given Pereboom the mistaken impression that my model is unable to successfully deal with white-collar crime. That, however, is not the case. There is no in principle reason why the public health model cannot be extended to include the social determinants of white-collar crime as well. As such, the public health-quarantine model is, in fact, well suited to prevent such crimes.

6 ELIZABETH SHAW

Elizabeth Shaw is another fellow free will skeptic and anti-retributivist. She was also an early exponent of the Epistemic Argument against retributivism, which I defend a version of in Chapter 3 of the book.[141] Like Derk Pereboom, Shaw and I generally agree on most things when it comes to free will (or the lack thereof) and punishment. In her comments here, however, she maintains that: 'I disagree with Caruso's suggestion that the epistemic argument is not applicable to the Public Health-Quarantine (PHQ) model. However, I argue that this does not undermine Caruso's overall theory, because it is plausible that his theory can satisfy the standard of credibility required by the epistemic argument'.[142] I welcome Shaw's critical comments and see them as providing additional support for my overall view. While Shaw disagrees with me about my use of the punitive/nonpunitive and intentional/nonintentional harming distinctions to exempt the PHQ model, we ultimately end up in the same place – i.e., concluding that the epistemic argument provides a powerful argument against the retributive justification of legal punishment *but not* the public health-quarantine model. I wish all my critics were so charitable. Despite our general agreement, though, there are still a few points of disagreement I would like to push back on.

6.1 The PHQ Model and the Standard of Proof

Shaw begins by correctly noting that the epistemic argument is based on the plausible idea that it is wrong to (intentionally) inflict serious harm on people, unless there are very strong grounds for believing that harming them is justified. She writes:

> there should be a presumption against (intentionally) inflicting serious harm, the burden of proof should lie with the person who is in favour of inflicting serious harm, and the

140. Pereboom, 'Non-free deterrence' (n 128).
141. E Shaw, *Free Will, Punishment, and Criminal Responsibility* (PhD thesis, University of Edinburgh 2014). See also Pereboom, *Living Without Free Will* (n 69); Pereboom, *Free Will, Agency, and Meaning in Life* (n 40); R Double, 'The Moral Hardness of Libertarians' (2002) 5 Philo 226; B Vilhauer, 'Free Will and Reasonable Doubt' (2009) 46 American Philosophical Quarterly 131; M Corrado, 'Punishment and Burden of Proof' (2017) UNC Legal Studies Research Paper; GD Caruso, 'Justice Without Retribution: An Epistemic Argument against Retributive Criminal Punishment' (2020) 13 Neuroethics 13; Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (n 1) Ch 3.
142. E Shaw, 'The Epistemic Argument against Retributivism' (2021) The Journal of Legal Philosophy (this issue).

> argument for inflicting serious harm must be established to a high standard of credibility. Like several other theorists who have applied this epistemic argument to punishment, Caruso persuasively argues that the appropriate standard of credibility for justification of punishment is 'beyond reasonable doubt'.

She goes on, however, to argue that, 'Caruso's case would be stronger if he modified his view about the non-applicability of the epistemic argument to his own PHQ model'.

Shaw's main point of disagreement is with my claim that, while proponents of retributive punishment must bear a heavy burden of proof, the public health-quarantine model need not be held to the beyond a reasonable doubt standard since it is nonpunitive and does not involve harming people intentionally. Against this claim, Shaw presents three main objections. First, she argues that, 'although the PHQ model can legitimately be regarded as nonpunitive, it should be held to the same epistemic standard as penal theories, because the PHQ model resembles punishment in certain relevant respects (and might be called quasi-punitive)'. Second, she argues that, 'the public health-quarantine model does involve inflicting harm intentionally'. Third, she points out that, 'Caruso cites unintentional harms when justifying applying a high epistemic standard of proof to other penal theories, so it would be inconsistent for him to argue that his theory should be exempt from that standard on the basis that his approach (supposedly) does not involve intentional harm'. Despite these objections, however, Shaw goes on to argue that, 'the public health-quarantine model meets the required standard of credibility, or, at least, that it could form a major component of an approach that would satisfy that standard'.

The last of Shaw's points is a welcome one since it maintains that, regardless of Shaw's objections, the public health-quarantine mode *can* meet the higher standard of credibility needed to justify the incapacitation of seriously dangerous criminals. In some ways, it makes Shaw's objections a form of 'inside baseball' where the disputing parties agree on the final outcome or conclusion but disagree on how best to reach it. But since I agree with Shaw that *how* one reaches a particular conclusion is just as important as the conclusion itself, I will now attempt to address some of her key critical points.

6.2 Replies to Shaw

With regard to the punitive/nonpunitive distinction, Shaw does not contest the claim that the public health-quarantine model 'can legitimately be classed as non-punitive'. Nevertheless, she maintains that, 'the public health-quarantine model clearly *resembles* punishment in terms of the seriousness of the harms imposed on offenders and the restrictions on the offender's liberty'. According to Shaw, '[t]hese similarities seem relevant to one of the key intuitions that motivate holding theories of punishment to a high standard of credibility – the intuition that seriously harming people requires strong justification'. While I agree with Shaw that the *methods* used in incapacitation resemble in some way, though importantly not others, those used in punishment, I think the real thing that sets the public health-quarantine model apart is that the *justification* it provides for incapacitation and other liberty-limiting restrictions appeals to the right of self-defence and defence of others and not a retributive justification that appeals to the assumption that wrongdoers are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done (appropriately qualified). Justifying retributive legal punishment is difficult because the claim that agents are morally responsible in the basic desert sense provided by both libertarian and compatibilist accounts faces powerful and unresolved objections and as a result falls far short of the high epistemic bar needed to justify such intentional harms. *Eliminative harming*, on the other hand – i.e., harming someone to eliminate a threat they pose – is much easier to justify. (More on this in a moment.)

Furthermore, while the liberty-limiting measures the public health-quarantine model permits may resemble punishment in some ways, we must not overlook the many ways it does not. Since legal punishment seeks to make wrongdoers suffer and requires the intentional imposition of a penalty for conduct that is represented as a violation of a law of the state, and since the public health-quarantine model does not involve punishment in this way, it offers a nonpunitive alternative to treatment of criminals. 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. The same is true when we incapacitate the criminally dangerous in order to protect society. The right of self-defence 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 behaviour 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 specifically for punitive purposes, we would need to completely redesign our institutions so that the physical environments and spaces we incapacitate people in better serve the goal of rehabilitation and reintegration.

Additionally, 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. For one, felony voter disenfranchisement policies would be ended, since they serve no forward-looking benefit and are ultimately a hindrance to the efforts of society to rehabilitate and reintegrate former felons. The sentence of life without parole would also be removed as an option, since it precludes from the outset the possibility of rehabilitation, violates the principle of least infringement, discourages the state from working to improve the well-being of offenders since they are seen as 'lost causes', and prevents the reassessment of individual cases as circumstances change. Three strikes laws would likewise need to be reversed since they prevent individual cases from being judged on their own terms, which would be needed if we are to accurately assess the threat an individual poses to society moving forward. Not all felonies are equal. And the fact that one person has committed three strikes does not mean they represent the same threat to society as another. Three strikes laws also often run afoul of the principle of least infringement.[143] In the book, I also endorse the *principle of normality*,[144] which maintains that offenders should have all the same rights as all other citizens and should be placed in the lowest

143. See Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (n 1) 25.
144. For an official statement of the principle, see: <www.kriminalomsorgen.no/information-in-english.265199.no.html>.

possible security regime. These are only a few examples of the many implications the public health-quarantine model would have, but hopefully they provide a sense of the wide-ranging nature of the reforms that would be required and the various ways a non-punitive approach would differ from traditional punishment.

I also maintain that there is an important difference between the state *intentionally harming* individuals (which is an essential component of punishment) and the state causing *unintentional harm* (which would apply to any harms caused by my nonpunitive system of incapacitation) – and it is arguable that the former must meet a higher epistemic bar. Shaw challenges this claim by arguing that: 'when the state intentionally imposes severe burdens on an offender, such as detaining him for a very long period in a secure facility, this constitutes harming him. Thus, one cannot claim that one intends to impose these severe burdens on the offender without intending to harm him'. In response, I would offer two replies. First, I would argue that while the distinction between *foreseeable-but-unintended harm* and *intended harm* is sometimes difficult to draw, we should not dismiss the importance of the distinction to the epistemic argument or moral issues in general. The *principle of double-effect*, for instance, maintains that if doing something morally good has a morally bad side-effect it's ethically permissible to do it provided the bad side-effect wasn't intended – and this is true even if we foresaw that the bad effect would probably happen.[145] Thomas Aquinas is credited with introducing the principle of double effect in his discussion of the permissibility of self-defence in this *Summa Theologica*.[146] Killing one's assailant is justified, he argued, provided one does not intend to kill him.

The principle of double effect is now widely accepted among moral theorists, no doubt because the example cited as illustrations of it have considerable intuitive appeal. Here are some examples drawn from the *Stanford Encyclopedia of Philosophy* entry on the principle:[147]

> The terrorist bomber aims to bring about civilian deaths in order to weaken the resolve of the enemy: when his bombs kill civilians this is a consequence that he intends. The tactical bomber aims at military targets while foreseeing that bombing such targets will cause civilian deaths. When his bombs kill civilians this is a foreseen but unintended consequence of his actions. Even if it is equally certain that the two bombers will cause the same number of civilian deaths, terrorist bombing is impermissible while tactical bombing is permissible.
>
> To kill a person whom you know to be plotting to kill you would be impermissible because it would be a case of intentional killing; however, to strike in self-defense against an aggressor is permissible, even if one foresees that the blow by which one defends oneself will be fatal.
>
> Sacrificing one's own life in order to save the lives of others can be distinguished from suicide by characterizing the agent's intention: a soldier who throws himself on a live grenade intends to shield others from its blast and merely foresees his own death; by contrast, a person who commits suicide intends to bring his or her own life to an end.

I maintain that in accordance with the principle of double effect, it is permissible, morally speaking, to restrict the liberty of those individuals who pose a significant threat to public health and safety, provided the harm caused by such restrictions is not intended. Furthermore, the permissibility of this kind of eliminative harming is

145. See A McIntyre, 'Doctrine of Double Effect', Stanford Encyclopedia of Philosophy (2018) available at: <https://plato.stanford.edu/entries/double-effect/>.
146. Summa Theologica (II-II, QU. 64, Art.7).
147. See McIntyre, 'Doctrine of Double Effect' (n 145).

significantly different than the kind of harming involved in retributive punishment – which *intentionally* seeks to impose harm or harsh treatment on wrongdoers simply because they deserve it.

This brings me to my second reply. Unlike retributive punishment, the justification of quarantine (in the case of disease) and incapacitation (in the case of dangerous criminals) is grounded in the right of self-defence and defence of others. Since this right in no way depends upon the questionable notions of free will and basic desert moral responsibility, any harms it causes those individuals who are quarantined or incapacitated would be justified on the grounds of self-defence and prevention of harm to others (even if such harms remain regrettable). The right of self-defence and defence of others justifies temporarily limiting one's liberty under certain constraints and restrictions, even if it does not justify punishment in the traditional sense. This means that *even if* we applied the epistemic argument to the public health-quarantine model itself, the burden of proof would *not* be on justifying the harms caused by quarantine or incapacitation – which is a form of eliminative harming that almost everyone agrees is permissible. Rather, the epistemic burden would be on establishing, to a high degree, an individual poses a *significant enough threat to public health and safety* to justify the right of self-defence. That is, in the case of the public health-quarantine model, the epistemic standard that would need to be satisfied would apply, *not* to the harms caused by intentional punishment (which is not involved in cases of eliminative harming) or the assumption that wrongdoers are free and morally responsible (which it denies). Instead, one would need to establish that an individual poses a significant enough threat to another's life, liberty or property to justify harming them on the grounds of self-defence, since it is almost universally agreed that in such circumstances eliminative harming would, in fact, be permitted. This is an extremely important point and one that is often overlooked by critics.

For the preceding reasons, I do not think the epistemic argument applies to the public health-quarantine model, at least not in the same way it applies to retributive legal punishment. While both might involve measures that cause harm to those they are imposed on, retributivism seeks to intentionally harm wrongdoers on the epistemically suspect assumption that individuals are free and morally responsible and hence justly deserve to suffer for the wrongs they have done, while the public health-quarantine model only seeks to eliminate the harm posed by an individual. The former, I maintain, requires a much higher burden of proof since it attempts to justify a set of punitive practices and policies that cause a great deal of intentional pain and suffering. The latter is easier to justify, I contend, since it involves only a kind of eliminative harming that seeks a morally good end and is therefore consistent with the principle of double effect. Furthermore, even if the public health-quarantine model was required to justify the harms it causes, the epistemic challenge would be to satisfy whatever burden we think is required to show that an individual poses a significant enough threat to another's life, liberty, or property to justify harming them on the grounds of self-defence. This is a significantly different challenge than the one facing retributivism, which requires what I claim is an insurmountable task – proving beyond a reasonable doubt that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done in a purely backward-looking, non-consequentialist sense.

The remainder of Shaw's paper is favourable to the public health-quarantine model since it argues that it would be able to meet, if required, the same standard of credibility demanded by the epistemic argument. In particular, she argues that 'decision-making under conditions of "moral uncertainty" would, firstly, favour some of the core claims of the PHQ model, and, secondly, would favour the *outcomes* concerning who should/should not be subjected to state coercion and concerning the severity of coercive measures that are implied by the PHQ model'. With regard to the first point, Shaw appeals to the same distinction I stressed above – i.e., the difference between the public health-quarantine model's justification for overriding the principle of harm avoidance, which appeals to the right of self-defence, and the justification provided by the retributivist, which appeals to the philosophically disputed notions of free will and basic desert moral responsibility. Shaw writes that, 'The PHQ model identifies a reason for overriding the harm avoidance principle that seems hard to dispute'. She goes on to argue:

> A core claim of the PHQ model is that society has a right, grounded in self-defence, to detain those whose criminal conduct demonstrates that they pose a serious threat to others, e.g., rapists and murderers. If a theorist, who had previously endorsed retributivism, came to believe that there is not enough certainty about the soundness of retributivism to allow retributivism to override the harm avoidance principle what should this theorist do? Surely, it would not be rational for such a theorist to say that, if retributivism can no longer provide a sufficient basis for interfering with the liberty of rapists and murderers, such offenders may not be interfered with at all. True, a retributivist, *qua* retributivist, would not take dangerousness to be part of the justification for interference with an offender's liberty. But it seems hard to deny that such a theorist, *qua* ethical theorist under conditions of moral uncertainty, would agree with Caruso that some interference with the offender's liberty would be justified.

I wholeheartedly endorse this argument and believe it helps additionally explain why the epistemic argument is successful against retributivism but not the public health-quarantine model.

Shaw's second point about outcomes also supports the public health-quarantine model. Here she writes:

> Endorsing a cautious approach when reasoning under conditions of moral uncertainty seems to imply that one should opt for a model of criminal justice which (compared with other theories) would recommend subjecting a relatively small number of people to state coercion and which would, overall, recommend relatively lenient responses to criminal behaviour. Caruso's theory would fit this description. Caruso provides a detailed defence of a range of robust safeguards against unjustified punishment. For example, his account stresses the importance of liberty, which he argues should only be infringed in accordance with various principles, including the 'principle of least infringement' … The cumulative effect of these strategies is that Caruso's theory is among the most lenient of mainstream approaches to dealing with criminal behaviour. Decision-makers under conditions of moral uncertainty would therefore have reasons to favour the outcomes implied by Caruso's theory, even if they disagreed about his rationale.

Again, I wholeheartedly agree with Shaw and am extremely grateful for her way of articulating this important point. When reasoning under conditions of moral uncertainty – i.e., when there is no consensus among theorists on whether agents are free and morally responsible in the relevant sense – I completely agree that we should prefer the option that produces the least amount of harm. Importantly, though, I would also add that under conditions of moral uncertainty we should *also* favour nonpunitive approaches to crime over punitive ones, which the public health-quarantine model also provides.

In conclusion, I endorse both of Shaw's points about decision-making under moral uncertainty and am willing to take them onboard. I am extremely gratefully to Shaw for both her insightful criticisms and her supportive ways of defending the public health-quarantine model. In the end, we both agree that the epistemic argument is successful against retributivism but not the public health-quarantine model. That, to me, is the important point, regardless of how one frames the victory – i.e., in terms of the ability of the public health-quarantine model to overcome the same epistemic burden as retributivism *or* to point to the important ways the epistemic burden differs in the case of eliminative harming.