Retributivism, Free Will, and the Public Health-Quarantine Model
Gregg D. Caruso

****

Within the criminal justice system one of the most prominent justifications for legal punishment, both historically and currently, is *retributivism*. The retributive justification of legal punishment maintains that, absent any excusing conditions, wrongdoers are morally responsible for their actions and *deserve* to be punished in proportion to their wrongdoing. Unlike theories of punishment that aim at deterrence, rehabilitation, or incapacitation, retributivism grounds punishment in the *blameworthiness* and *desert* of offenders. It holds that punishing wrongdoers is intrinsically good. For the retributivist, wrongdoers deserve a punitive response proportional to their wrongdoing, even if their punishment serves no further purpose. This means that the retributivist position is not reducible to consequentialist considerations nor in justifying punishment does it appeal to wider goods such as the safety of society or the moral improvement of those being punished.

In this chapter, I outline six distinct reasons for rejecting retributivism, not the least of which is that it's unclear that agents possess the kind of free will and moral responsibility needed to justify it. I then sketch my novel non-retributive alternative, which I call the *public health-quarantine model*. As we'll see, the model draws on the public health framework and prioritizes prevention and social justice. I contend that it not only offers a stark contrast to retributivism, it also provides a more humane, holistic, and effective approach to dealing with criminal behavior, one that is superior to both retributivism and other leading non-retributive alternatives. I begin by taking a closer look at the retributive justification of punishment.

1. Retributivism

According to the retributivist justification of legal punishment, wrongdoers deserve the imposition of a penalty solely for the backward-looking reason that they have knowingly done wrong. Michael S. Moore, a leading retributivist, highlights this purely backward-looking nature of retributivism when he writes:

> [R]etributivism is the view that we ought to punish offenders because, and only because, they deserve to be punished. Punishment is justified, for a retributivist, solely by the fact that those receiving it deserve it. Punishment may deter future crime, incapacitate dangerous persons, educate citizens in the behaviour required for a civilized society, reinforce social cohesion, prevent vigilante behaviour, make victims of crime feel better, or satisfy the vengeful desires of citizens who are not themselves crime victims. Yet for the retributivist these are a happy surplus that punishment produces and form no part of what makes punishment just: for a retributivist, deserving offenders should be punished

even if the punishment produces none of these other, surplus goof effects. (1997: 153; see also 1987, 1993)

This backward-looking focus on desert is a central feature of all traditional retributive accounts of punishment (see, e.g., Kant 1790; von Hirsch 1976, 1981, 2007, 2017; Husak 2000; Kershnar 2000, 2001; Berman 2008, 2011, 2013, 2016; Walen 2014). And it is important to emphasize that the desert invoked in retributivism (in the classical or strict sense) is *basic* in the sense that it is not in turn grounded in forward-looking reasons such as securing the safety of society or the moral improvement of criminals. Thus, for the retributivist, the claim that persons are morally responsible for their actions in the *basic desert* sense is crucial to the state's justification for giving them their just deserts in the form of punishment for violations of the state's laws.[1]

In the US criminal justice system, the retributivist justification of legal punishment and the attendant proportionality requirement are widely embraced. In fact, a number of sentencing guidelines in the United States have adopted the retributivist conception of desert as their core principle,[2] and it is increasingly given deference in the "Purposes" section of state criminal codes,[3] where it can be the guiding principle in the interpretation and application of the code's provisions.[4] Indeed, the American Law Institute recently revised the Model Penal Code so as to set desert as the official dominant principle for sentencing.[5] And courts have identified desert as the guiding principle in a variety of contexts,[6] as with the Supreme Court's enthroning of retributivism as the "primary justification for the death penalty"[7] (Robinson 2008: 145–146). Additional examples can be found in legislation, judicial decisions, sentencing guidelines, and criminal codes in England, Wales, Scotland, Australia, Canada, New Zealand, and Israel (see, e.g., Dingwall 2008; von Hirsch 2017).

Depending on how retributivists view the relationship between desert and punishment, we can identify three different varieties of the view – *weak*, *moderate*, and *strong*.[8] *Weak retributivism* maintains that negative desert, which is what the criminal law is concerned with when it holds wrongdoers accountable,[9] is merely necessary but not sufficient for punishment. That is, weak

[1] Retributivists typically also hold, in addition, that just punishments must be *proportional* to wrongdoing. Both the justificatory thesis and the proportionality requirement for punishments are reflected in Mitchell Berman's statement of retributivism: "A person who unjustifiably and inexcusably causes or risks harm to others or to significant social interests deserves to suffer for that choice, and he deserves to suffer in proportion to the extent to which his regard or concern for others falls short of what is properly demanded of him" (2008: 269).

[2] E.g., 204 Pa. Code Sect. 303.11 (2005); see also (Tonry 2004).

[3] E.g., Cal. Penal Code Sect. 1170(a)(1) (West 1985): "The legislature finds and declares that the purpose of imprisonment for crime is punishment."

[4] E.g., Model Penal Code Sect. 1.02(2) (Official Draft 1962).

[5] American Law Institute, Model Penal Code Sect. 1.02(2) adopted May 24, 2017.

[6] See, e.g., the US cases *Spaziano v. Florida*, 468 U.S. 447, 462 (1984); *Gregg v. Georgia*, 428 U.S. 153, 183-184 (1976); Cotton (2000).

[7] *Spaziano v. Florida*, 468 U.S. at 461.

[8] See, e.g., Alexander, Ferzan, and Morse (2009: 7-10), Walen (2014).

[9] *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

retributivism maintains that while desert is a necessary condition for punishment, it is not enough on its own to justify punishment – other conditions must also be met. As Alec Walen describes it, weak retributivism is the view that "wrongdoers forfeit their right not to suffer proportional punishment, but that the positive reasons for punishment must appeal to some other goods that punishment achieves, such as deterrence or incapacitation" (2014). Wrongdoing, on this view, is merely a necessary condition for punishment: "The desert of the wrongdoer provides neither a sufficient condition for nor even a positive reason to punish" (Walen 2014; see also Mabbott 1939; Quinton 1954).

Moderate retributivism, on the other hand, maintains that negative desert is necessary and sufficient for punishment but that desert does not mandate punishment or provide an obligation to punish in all circumstances – that is, there may be other goods that outweigh punishing the deserving or giving them their just deserts (Robinson and Cahill 2006). Leo Zaibert, while eschewing the taxonomy offered here, defends a kind of moderate retributivism when he argues:

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

Mitchell Berman also defends a form of moderate retributivism, which he calls "modest retributivism" (2016), since he maintains that negative desert grounds a justified reason to punish, but not a duty. For moderate retributivists, negative desert is sufficient to justify punishment but other values and considerations may outweigh inflicting the deserved punishment.

Lastly, *strong retributivism* maintains that desert is necessary and sufficient for punishment but it also grounds a duty to punish wrongdoers. Immanuel Kant is perhaps the most famous representative of this latter view, since he famously argued that the death penalty was not only deserved but also obligatory in cases of murder:

> [W]hoever has committed murder, must die. There is, in this case, no juridical substitute or surrogate, that can be given or taken for the satisfaction of justice. There is no Likeness or proportion between Life, however painful, and Death; and therefore there is no Equality between the crime of Murder and the retaliation of it but what is judicially accomplished by the execution of the Criminal. (1790: Part II: 6)

---

punishing those not guilty of wrongdoing (who deserve no punishment), and punishing the guilty more than they deserve (i.e., inflicting disproportional punishment)" (2014, s. 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.

He goes on to write:

> Even if a civil society resolved to dissolve itself with the consent of all its members – as might be supposed in the case of a people inhabiting an island resolving to separate and scatter themselves throughout the whole world – the last murderer lying in prison ought to be executed before the resolution was carried out. This ought to be done in order that every one may realize the desert of his deeds, and that blood-guiltiness may not remain upon the people; for otherwise they might all be regarded as participators in the murder as a public violation of justice. (1790: Part II: 6)

Of course, not all retributivists support the death penalty – in fact, many contemporary retributivists do not – but in the above quote, Kant is embodying the strong retributivist view that we are not only justified in giving offenders their just deserts by punishing them, we have a duty to do so. Michael S. Moore also defends a form of strong retributivism and argues, like Kant, that society has a duty to punish culpable offenders:

> We are justified in punishing because and only because offenders deserve it. Moral responsibility ("desert") in such a view is not only necessary for justified punishment, it is also sufficient. Such sufficiency of justification gives society more than merely a right to punish culpable offenders. It does this, making it not unfair to punish them, but retributivism justifies more than this. For a retributivist, the moral responsibility of an offender also gives society the duty to punish. Retributivism, in other words, is truly a theory of justice such that, if it is true, we have an obligation to set up institutions so that retribution is achieved. (1997: 91)

Strong retributivists therefore defend two distinct claims: (1) that negative desert is sufficient to justify punishing wrongdoers on the grounds that they deserve it and (2) that we have a duty to do so. Moderate retributivists, on the other hand, seek only to defend the first claim.

In what follows, I will limit my discussion to moderate and strong varieties of retributivism and leave weak retributivism aside. I will do so because, first, most leading retributivists defend one of these stronger forms of retributivism and it is my desire to address the dominant view, not a subordinate view held by few – see, for example, Moore (1997, 1987, 1993), Kershnar (2000, 2001), Husak (2000), Berman (2008, 2011), von Hirsch (1976, 2007, 2017), Alexander (2013), Alexander, Ferzan, and Morse (2009). Second, 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" (Alexander, Ferzan, and Morse 2009: 7). 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" (2008: 86; emphasis added). 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. (2016)

And Alec Walen in his *Stanford Encyclopedia of Philosophy* entry on "Retributive Justice" (2014) 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 impermissible to intentionally punish the innocent or to inflict disproportionately large punishments on wrongdoers.

Lastly, 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 makes this point rather clear. For these reasons, I will take as my 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. Since this core claim is held in common among all moderate and strong varieties of retributivism, I will henceforth drop the moderate/strong distinction and focus instead on this shared feature.

2. Rejecting Retributivism

While retributivism provides one of the main sources of justification for punishment within criminal justice systems, there are good philosophical and practical reasons for rejecting it. One such reason is that it is unclear that agents *truly deserve* to suffer for the wrongs they have done in the sense required by retributivism. This is because, for an agent to *deserve* to suffer for the wrongs they have done in the purely backward-looking sense required for retributivism, they would need to possess the kind of control in action—i.e., free will—required for basic desert moral responsibility (see Caruso 2021a). Yet there are good philosophical reasons for thinking agents are never free and morally responsible in this sense. *Hard determinists*, for example, have long argued that determinism is true and incompatible with free will and basic desert moral responsibility—either because it precludes the *ability to do otherwise* (leeway incompatibilism) or because it is inconsistent with one's being the "ultimate source" of action (source incompatibilism). More recently, a number of contemporary philosophers have presented additional arguments against basic desert moral responsibility that are agnostic about determinism (e.g., Pereboom 2001, 2014; Strawson 1986; Levy 2011; Waller 2011; Caruso 2012, 2021a; Dennett and Caruso 2021).

In my own work, I have offered two distinct sets of arguments in support of *free will skepticism* (see Caruso 2012, 2021a; Dennett and Caruso 2021). The first features distinct arguments that target the three leading rival views—event-causal libertarianism, agent-causal libertarianism, and compatibilism—and then claims the skeptical position is the only defensible position that remains standing. It's a form of *hard incompatibilism*, which maintains that free will is incompatible with *both* causal determination by factors beyond the agent's control *and* with the kind of indeterminacy in action required by the most plausible versions of libertarianism. Against the view that free will is compatible with the causal determination of our actions by natural

factors beyond our control, I argue that there is no relevant difference between this prospect and our actions being causally determined by manipulators (see Pereboom 2001, 2014; Caruso 2012, 2021a). Against event causal libertarianism, I object (among other things) that on such accounts agents are left unable to settle whether a decision occurs and hence cannot have the control required for moral responsibility (see Pereboom 2001, 2014l Caruso 2012, 2021a). I further maintain that non-causal accounts of free will suffer from the same problem (see Pereboom 2001, 2014). While agent-causal libertarianism could, in theory, supply this sort of control, I argue that it cannot be reconciled with our best philosophical and scientific theories about the world (see Pereboom 2001, 2014; Caruso 2012, 2021a) and faces additional problems accounting for mental causation (Caruso 2012, 2021a). Since this exhausts the options for views on which we have the sort of free will at issue, I conclude that free will skepticism is the only remaining position.

In addition to hard incompatibilism, I also defend a second, independent argument against free will which maintains that regardless of the causal structure of the universe, free will and basic desert moral responsibility are incompatible with the pervasiveness of *luck*—a view sometimes called *hard luck.* This argument is intended not only as an objection to libertarianism but extends to compatibilism as well. At the heart of the argument is the following dilemma, which Neil Levy (2011) calls the *luck pincer*: Either actions are subject to *present luck* (luck around the time of action), or they are subject to *constitutive luck* (luck in who one is and what character traits and predispositions one has), or both. Either way, luck undermines free will and basic desert moral responsibility since it undermines responsibility-level control.

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." 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.

What these and other arguments for free will skepticism have in common, and what they share with classical hard determinism, is the thesis that what we do and the way we are is ultimately the result of factors beyond our control and because of this we are never morally responsible for our actions in the basic desert sense—the sense required for retributive punishment. This is not to say, of course, that other conceptions of responsibility cannot be reconciled with determinism, chance, or luck. Nor is it to deny that there may be good reasons to maintain certain systems of moral protest in the face of bad behavior (see Pereboom 2014, 2021). Rather, it is to insist that to hold people *truly* or *ultimately* morally responsible for their actions—i.e., to hold them responsible in a non-consequentialist desert-based sense—would be to hold them responsible for

the results of the morally arbitrary, for what is ultimately beyond their control which is fundamentally unfair and unjust.

My first argument, then, against retributivism—i.e., the *Skeptical Argument*—maintains that free will skepticism undermines the retributive justification for punishment since it does away with the idea of *basic desert*. The justification of retributivism depends on the assumption that criminals are (or at least can be) deserving of blame in the basic desert sense for their criminal behavior. But, if free will skepticism is true, then *no one* is ever deserving of blame in the basic desert sense for any of their actions. So, the truth of free will skepticism entails that retributive punishment cannot be justified, and thus retributivism should be rejected.

But what if one is not totally convinced by the arguments for free will skepticism? Well, I maintain that *even in the face of uncertainty about the existence of free will*, it remains unclear whether retributive punishment is justified. This is because the burden of proof lies on those who want to inflict harm on others to provide good justification for such harm. This means that retributivist who want to justify legal punishment on the assumption that agents are free and morally responsible (and hence *justly deserve* to suffer for the wrongs they have done) must justify that assumption. And they must justify that assumption in a way that meets a high epistemic standard of proof since the harms caused in the case of legal punishment are often quite serve. It is not enough to simply point to the mere possibility that agents possess libertarian or compatibilist free will. Nor is it enough to say that the skeptic arguments against free will and basic desert moral responsibility fail to be conclusive. Rather, a positive and convincing case must be made that agents are in fact morally responsible in the basic desert sense, since it is the backward-looking desert of agents that retributivists take to justify the harm caused by legal punishment.

This brings me to my second argument against retributivism, the so-called *Epistemic Argument* (Caruso 2020, 2021a). Versions of this argument have been developed by Derk Pereboom (2001; Benjamin Vilhauer (2009a, 2012, 2015), Elizabeth Shaw (2014, 2021), Michael Corrado (2017), and Sofia Jeppsson (2020), but my version of the argument can be summarized as follows:

(1) Legal punishment intentionally inflicts harms on individuals and the justification for such harms must meet a high epistemic standard. If it is significantly probable that one's justification for harming another is unsound, then, prima facie, that behavior is seriously wrong.

(2) The retributive justification for legal punishment assumes that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done in a backward-looking, non-consequentialist sense (appropriately qualified and under the constraint of proportionality).

(3) If the justification for the assumption that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done does not meet the high epistemic standard specified in (1), then retributive legal punishment is prima facie seriously wrong.

(4) The justification for the claim that agents are morally responsible in the basic desert sense provided by both libertarians and compatibilists face powerful and unresolved objections and as a result fall far short of the high epistemic bar needed to justify such harms.

(5) Hence, retributive legal punishment is unjustified and the harms it causes are prima facie seriously wrong.

Note that the Epistemic Argument requires only a weaker notion of skepticism than the one defended in the Skeptical Argument, namely one that holds that the justification for believing that agents are morally responsible in the basic desert sense, and hence justly deserve to suffer for the wrongs they have done, is too weak to justify the intentional suffering caused by retributive legal punishment.

Premise (1) places the burden of proof on those who want to justify legal punishment, since the harms caused in this case are often quite severe—including the loss of liberty, deprivation, and in some cases even death. As Victor Tadros spells out these harms:

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

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

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

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

Our ordinary everyday practices also place the burden of proof on those who knowingly and intentionally cause harm to others. In fact, even in cases where harm is *foreseeable* but not intended, we often demand a high level of justification. Let us say a newspaper receives a tip on a story that will likely cause great harm to a public figure, potentially sinking their career. In

such circumstances, good journalistic standards demand that the story be independently verified and properly vetted before it is run. If the newspaper were to run the story without properly vetting it, and later discover that the tip came from an organization who seeks to undermine the public's trust in the media, we would rightly condemn the newspaper for not applying a higher epistemic standard. Things are even clearer when the harm caused is intentional, like in the case of a just war or when a nation decides to use deadly force.

In the case of legal punishment where the severity of harm is beyond question, I maintain that we should place the highest burden possible upon the state. If the state is going to punish someone for first-degree murder, say, then the epistemic bar that needs to be reached is guilt beyond a reasonable doubt. But does this burden of proof carry over to theoretical debates—for example, the debate over free will and moral responsibility? Here I follow Pigliucci and Maarten (2014) as well as Corrado (2017: 3) in distinguishing between *evidential* burden of proof, which comes into play only when there is no costs associated with a wrong answer, and *prudential* burden of proof, which comes into play precisely when there are significant costs associated with a wrong answer. As Corrado applies the distinction to theoretical matters:

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

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

Benjamin Vilhauer, for instance, has persuasively argued that "[i]f it can be reasonably doubted that someone had free will with respect to some action, then it is a requirement of justice to refrain from doing serious retributive harm to him in response to that action" (2009: 131). Derk Pereboom has also proposed applying the reasonable doubt standard:

> Punishment – in particular, punishment designed to satisfy the retributive goals – harms people. If one aims to harm another, the justification must meet a high epistemic standard. If it not beyond reasonable doubt that retributivist justifications are disguised vengeful justification, and vengeful justification are illegitimate, then there is reason to believe that it is immoral to justify punishment policy retributivistically. More generally, where there is a substantial likelihood that one's justification for harming someone is illegitimate, then harming that person on the basis of that justification could well be morally wrong. (2001: 161)

The proof-beyond-a-reasonable-doubt standard is the appropriate epistemic standard to apply when we are talking about intentional harm and institutional punishment. When the stakes are high, as they are with legal punishment, both the law and everyday practice demand that we set the epistemic bar accordingly. As Vilhauer notes, the prudential burden of proof beyond a reasonable doubt has a close kinship to another "reasonable doubt" principle, which is widely recognized to be a requirement of justice: "[T]hat is the requirement in Anglo-American criminal legal proceedings that the accused can only be convicted of a crime if it is proven beyond reasonable doubt that he acted criminally" (2009: 133). The grounds for accepting this high epistemic standard for criminal conviction are the same as the grounds for accepting it with regard to premise (1).

When premise (1) is combined with (2), which is simply a statement of the retributivist justification for legal punishment, we get the requirement that retributivists must justify their core assumption – that is, that agents are free and morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done. As Vilhauer puts it:

> When the claim that someone has free will plays a role in a retributive justification of serious harm, that claim must be held to the same standard [as criminal conviction], for the same reason. That is, in this context, the claim that someone has free will plays a role in an argument for seriously harming someone, just as the claim that someone has committed a crime typically does. For this reason, it must be held to the "reasonable doubt" standard, just as the claim that someone has committed a crime must be. (2009: 134)

While this demand for justification is reasonable given the strength of (1), many retributivists simply deny or ignore it (see Caruso 2012 for a discussion of such views). And those libertarian and compatibilist accounts that do try to justify the assumption of free will, fail to overcome the high epistemic burden of proof needed to justify retributive harm. This is because they tend to be either scientifically implausible (as in the case of agent causation), empirically unwarranted (as in the case of event causal libertarianism), beg the question (as in the case of Strawson and other forms of compatibilism), or end up "changing the subject" (as in the case of Dennett and others).[10] Furthermore, the debate over free will has been waging for over two thousand years and reasonable people still disagree. In the face of such professional disagreement and uncertainty, I maintain that we should refrain from intentionally harming wrongdoers on the philosophically questionable assumption that they deserve it and that such punishment is intrinsically good.

In *Rejecting Retributivism* (2021a), I also develop four additional arguments against retributivism that are independent of worries over free will and basic desert moral responsibility. They include the *Misalignment Argument*, which maintains that it is philosophically problematic to impart to the state the function of intentionally harming wrongdoers in accordance with desert since it's not at all clear that the state is capable of properly tracking the desert and blameworthiness of individuals in any reliable way. This is because criminal law is not properly designed to account for all the various factors that affect blameworthiness, and as a result the *moral criteria of blameworthiness* is often misaligned with the *legal criteria of guilt* (see also

---

[10] See Caruso (2020, 2021a) for further details.

Kelly 2018). I also present a closely related argument, which I call *Poor Epistemic Position Argument (PEPA)*. It argues that for the state to be able to justly distribute legal punishment in accordance with desert, it needs to be in the proper epistemic position to know what an agent basically deserves, but since the state is (almost) never in the proper epistemic position to know what an agent basically deserves, it follows that the state is not able to justly distribute legal punishment in accordance with desert.

My final two arguments against retributivism are the *Indeterminacy in Judgment Argument* and the *Limited Effectiveness Argument*. The former maintains that how the state goes about judging the gravity of wrong done, on the one hand, and what counts as proportional punishment for that wrong, on the other, is wide open to subjective and cultural biases and prejudices, and as a result, the principle of proportionality in *actual practice* does not provide the kind of protections against abuse it promises. The latter argues that there are good additional pragmatic reasons for rejecting retributivism since it has limited effectiveness in promoting important social goals such as rehabilitation and reforming offenders. For more on these arguments, see Caruso (2021a).

I maintain that these six arguments—the Skeptical Argument, Epistemic Argument, Misalignment Argument, Poor Epistemic Position Argument, Indeterminacy in Judgment Argument, and Limited Effectiveness Argument—give us more than ample reason to reject retributivism (see Caruso 2021a).

3. The Public Health-Quarantine Model

If, then, we come to doubt or deny the existence of free will, or reject retributivism for other reasons, where does that leave us with regard to criminal justice? Many worry that without the justification of retributivism and the putative protection afforded by the principle of proportionality, we would be unable to successfully deal with criminal behavior. I contend, however, that this is not the case and that there is an ethically defensible and practically workable alternative to retributive legal punishment, one that is consistent with free will skepticism and preferable to other nonretributive alternatives. I call it the *public health-quarantine model* (see Caruso 2016, 2021a, 2021b; Pereboom and Caruso 2018; Caruso and Pereboom 2020).[11] The model not only provides a justification for the incapacitation of dangerous criminals consistent with free will skepticism, it also provides a broader and more comprehensive approach to criminal behavior generally since it draws on the public health framework and prioritizes prevention and social justice.

The public health-quarantine model is based on an analogy with quarantine and draws on a comparison between treatment of dangerous criminals and treatment of carriers of dangerous diseases. It takes as its starting point Derk Pereboom's famous account (2001, 2013, 2014). In its simplest form, it can be stated as follows: (1) Free will skepticism maintains that criminals are not morally responsible for their actions in the basic desert sense; (2) plainly, many carriers of

---

[11] I should note that there are other alternatives to classical retributivism that are consistent with the rejection of retributivism—such as consequentialist deterrence theories, educational theories, communicative theories, and mixed accounts. But I have elsewhere argued that these approaches have ethical problems of their own that are difficult to overcome and make them less desirable than my non-retributive and non-punitive alternative (see Caruso 2021).

dangerous diseases are not responsible in this or in any other sense for having contracted these diseases; (3) yet, we generally agree that it is sometimes permissible to quarantine them, and the justification for doing so is the right to self-protection and the prevention of harm to others; (4) for similar reasons, even if a dangerous criminal is not morally responsible for his crimes in the basic desert sense (perhaps because no one is ever in this way morally responsible) it could be *as* legitimate to preventatively detain him as to quarantine the non-responsible carrier of a serious communicable disease (Pereboom 2014: 156).

The first thing to note about the theory is that although one might justify quarantine (in the case of disease) and incapacitation (in the case of dangerous criminals) on purely utilitarian or consequentialist grounds, both Pereboom and I want to resist this strategy (see Pereboom and Caruso 2018; see also Caruso 2021a). Instead, on our view incapacitation of the dangerous is justified on the ground of the right to harm in self-defense and defense of others. That we have this right has broad appeal, much broader than utilitarianism or consequentialism has. In addition, this makes the view more resilient to a number of objections (see Caruso 2021a; Pereboom and Caruso 2018).

Second, the quarantine model places several constraints on the treatment of criminals (see Pereboom 2001, 2014; Pereboom and Caruso 2018; Caruso 2021a). First, as less dangerous diseases justify only preventative measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate restraints (Pereboom 2014: 156). In fact, for certain minor crimes perhaps only some degree of monitoring could be defended. Secondly, the incapacitation account that results from this analogy demands a degree of concern for the rehabilitation and well-being of the criminal that would alter much of current practice. Just as fairness recommends that we seek to cure the diseased we quarantine, so fairness would counsel that we attempt to rehabilitate the criminals we detain (Pereboom 2014: 156). If a criminal cannot be rehabilitated, however, and our safety requires his indefinite confinement, this account provides no justification for making his life more miserable than would be required to guard against the danger he poses (Pereboom 2014: 156).

Third, this account also provides a more resilient proposal for justifying criminal sanctions than other non-retributive options. One advantage it has, say, over consequentialist deterrence theories is that it has more restrictions placed on it with regard to using people merely as a means. For instance, as it is illegitimate to treat carriers of a disease more harmfully than is necessary to neutralize the danger they pose, treating those with violent criminal tendencies more harshly than is required to protect society will be illegitimate as well. In fact, in all our writings on the subject, Pereboom and I have always maintained the *principle of least infringement*, which holds that the least restrictive measures should be taken to protect public health and safety (Caruso 2016, 2017a, 2021a; Pereboom and Caruso 2018). This ensures that criminal sanctions will be proportionate to the danger posed by an individual, and any sanctions that exceed this upper bound will be unjustified.

In addition to these restrictions on harsh and unnecessary treatment, the model also advocates for a broader approach to criminal behavior that moves beyond the narrow focus on sanctions. On the model I have developed, the quarantine analogy is placed within the broader justificatory framework of *public health ethics* (Caruso 2016, 2017a, 2021a). Public health ethics not only

justifies quarantining carriers of infectious diseases on the grounds that it is necessary to protect public health, it also requires that we take active steps to *prevent* such outbreaks from occurring in the first place. Quarantine is only needed when the public health system fails in its primary function. Since no system is perfect, quarantine will likely be needed for the foreseeable future, but it should *not* be the primary means of dealing with public health. The analogous claim holds for incapacitation. Taking a public health approach to criminal behavior would allow us to justify the incapacitation of dangerous criminals when needed, but it would also make prevention a *primary function* of the criminal justice system. So instead of myopically focusing on punishment, the public health-quarantine model shifts the focus to identifying and addressing the systemic causes of crime, such as poverty, low social economic status, systematic disadvantage, mental illness, homelessness, educational inequity, abuse, and addiction (see Caruso 2021a).

In my recent *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (2021a), I argue that the social determinants of health (SDH) and the social determinants of criminal behavior (SDCB) are broadly similar, and that we should adopt a broad public health approach for identifying and taking action on these shared social determinants. I focus on how social inequities and systemic injustices affect health outcomes and criminal behavior, how poverty affects brain development, how offenders often have pre-existing medical conditions (especially mental health issues), how homelessness and education affects health and safety outcomes, how environmental health is important to both public health and safety, how involvement in the criminal justice system itself can lead to or worsen health and cognitive problems, and how a public health approach can be successfully applied within the criminal justice system. I argue that, just as it is important to identify and take action on the SDH if we want to improve health outcomes, it is equally important to identify and address the SDCB. And I conclude by offering eight broad public policy proposals for implementing a public health approach aimed at addressing the SDH and SDCB.

Furthermore, the public health framework I adopt sees *social justice* as a foundational cornerstone to public health and safety (Caruso 2016, 2021a). In public health ethics, a failure on the part of public health institutions to ensure the social conditions necessary to achieve a sufficient level of health is considered a grave injustice. An important task of public health ethics, then, is to identify which inequalities in health are the most egregious and thus which should be given the highest priority in public health policy and practice. The public health approach to criminal behavior likewise maintains that a core moral function of the criminal justice system is to identify and remedy social and economic inequalities responsible for crime. Just as public health is negatively affected by poverty, racism, and systematic inequality, so too is public safety. This broader approach to criminal justice therefore places issues of social justice at the forefront. It sees racism, sexism, poverty, and systemic disadvantage as serious threats to public safety and it prioritizes the reduction of such inequalities (see Caruso 2021a).

While there are different ways of understanding *social justice* and different philosophical accounts of what a theory of justice aims to achieve, I favor a *capability approach* according to which the development of capabilities—what each individual is able to do or be—is essential to human well-being (e.g., Sen 1985, 1999; Nussbaum 2011; Power and Faden 2006). For capability theorists, human well-being is the proper end of a theory of justice. And on the particular capability approach I favor, social justice is grounded in six key features of human

well-being: *health*, *reasoning*, *self-determination*, *attachment*, *personal security*, and *respect* (see Powers and Faden 2006; Caruso 2021a).[12] Following Powers and Faden (2006), I maintain that each of these six dimensions is an essential feature of well-being such that "a life substantially lacking in any one is a life seriously deficient in what it is reasonable for anyone to want, whatever else they want" (Powers and Faden 2006: 8). The job of justice is therefore to achieve a sufficiency of these six essential dimensions of human well-being, since each is a separate indicator of a decent life.

The key idea of capability approaches is that social arrangements should aim to expand people's capabilities—their freedom to promote or achieve *functionings* that are important to them. *Functionings* are defined as the valuable activities and states that make up human well-being, such as having a healthy body, being safe, or having a job. While they are related to goods and income, they are instead described in terms of what a person is able to do or be as a result. For example, when a person's need for food (a commodity) is met, they enjoy the functioning of being well-nourished. Examples of functionings include being mobile, being healthy, being adequately nourished, and being educated. The genuine opportunity to achieve a particular functioning is called a *capability*. *Capabilities* are "the alternative combination of functionings that are feasible for [a person] to achieve"—they are "the substantive freedom" a person has "to lead the kind of life he or she has reason to value" (Sen 1999: 87).

As Tabandeh, Gardoni, and Murphy describe:

> Genuine opportunities and actual achievements are influenced by what individuals have and what they can do with what they have. What they can do with what they have is a function of the structure of social, legal, economic, and political institutions and of the characteristics of the built-environment (i.e., infrastructure). For example, consider the functioning of being mobile. The number of times an individual travels per week can be an indicator of mobility achievement. When explaining a given individual's achievement or lack of achievement, a capability approach takes into consideration the conditions that must be in place for an individual to be mobile. For instance, the possession of certain resources, like a bike, may influence mobility. However, possessing a bike may not be sufficient to guarantee mobility. If the individual has physical disabilities, then the bike will be of no help to travel. Similarly, if there are no paved roads or if societal culture imposes a norm that women are not allowed to ride a bike, then it will become difficult or even impossible to travel by means of a bike. As this example makes clear, different factors will influence the number of times the individual travels. (Tabandeh, Gardoni, and Murphy 2017)

Thinking in terms of capabilities therefore raises a wider range of issues than simply looking at the amount of resources or commodities people have, because people have different needs. In the example given above, just providing bicycles to people will not be enough to increase the functioning of being mobile if you are disabled or prohibited from riding because of sexist social norms. A capabilities approach to social justice therefore requires that we consider and address a larger set of social issues.

---

[12] Note that this is a pared down list from the ones offered by Martha Nussbaum and other capability theorists (see Nussbaum 2011).

Bringing everything together, my public health-quarantine model characterizes the moral foundation of public health as social justice, not just the advancement of good health outcomes. That is, while promoting social goods (like health) is one area of concern, public health ethics as I conceive it is embedded within a broader commitment to secure a sufficient level of health and safety for all and to narrow unjust inequalities (see Powers and Faden 2006). More specifically, I see the capability approach to social justice as the proper moral foundation of public health ethics. This means that the broader commitment of public health should be the achievement of those capabilities needed to secure a sufficient level of human well-being—including, but not limited to, health, reasoning, self-determination, attachment, personal security, and respect. By placing social justice at the foundation of the public health approach, the realms of criminal justice and social justice are brought closer together. I see this as a virtue of the theory since it is hard to see how we can adequately deal with criminal justice without simultaneously addressing issues of social justice. Retributivists tend to disagree since they approach criminal justice as an issue of individual responsibility and desert, not as an issue of prevention and public safety. I believe it is a mistake to hold that the criteria of individual accountability can be settled apart from considerations of social justice and the social determinants of criminal behavior. Making social justice foundational, as my public health-quarantine model does, places on us a collective responsibility—which is forward-looking and perfectly consistent with free will skepticism—to redress unjust inequalities and to advance collective aims and priorities such as public health and safety. The capability approach and the public health approach therefore fit nicely together. Both maintain that poor health and safety are often the byproducts of social inequities, and both attempt to identify and address these social inequities in order to achieve a sufficient level of health and safety.

Summarizing the public health-quarantine model, then, the core idea is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. The resulting account would not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. The model also specifies attention to the well-being of criminals, which would change much of current policy. Furthermore, the public health component of the theory prioritizes prevention and social justice and aims at identifying and taking action on the social determinants of health and criminal behavior. This combined approach to dealing with criminal behavior, I maintain, is sufficient for dealing with dangerous criminals, leads to a more humane and effective social policy, and is actually preferable to the harsh and often excessive forms of punishment that typically come with retributivism.[13]

References
Alexander, Larry. 2013. "You Got What You Deserved." *Criminal Law and Philosophy* 7(2): 309-319.

Alexander, Larry, Kimberly Kessler Ferzan, and Stephen Morse. 2009. *Crime and Culpability: A Theory of Criminal Law*. New York: Cambridge University Press.

---

[13] For my replies to various objections to the public health-quarantine model, see Caruso 2021a, 2021b, 2021c; and Pereboom and Caruso 2018.

Berman, Mitchell. 2008. "Punishment and Justification." *Ethics* 118(2): 258-290.

Berman, Mitchell. 2011. "Two Kinds of Retributivism." In *Philosophical Foundations of Criminal Law*, edited by R.A. Duff and Stuart Green, 433-457. New York: Oxford University Press.

Berman, Mitchell. 2013. "Rehabilitating Retributivism." *Law and Philosophy* 32: 83-108.

Berman, Mitchell. 2016. "Modest Retributivism." In *Legal, Moral, and Metaphysical Truths: The Philosophy of Michael S. Moore*, edited by Kimberly Kessler Ferzan and Stephen J. Morse, 35-48. New York: Oxford University Press.

Boonin, David. 2008. *The Problem of Punishment.* New York: Cambridge University Press.

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

Caruso, Gregg D. 2016. "Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model." *Southwest Philosophical Review* 32(1): 25-48.

Caruso, Gregg D. 2020. "Justice with Retribution: An Epistemic Argument Against Retributive Criminal Punishment." *Neuroethics* 13(1): 13-28.

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

Caruso, Gregg D. 2021b. "Retributivism, Free Will Skepticism, and the Public Health-Quarantine Model: Replies to Kennedy, Walen, Corrado, Sifferd, Pereboom, and Shaw." *Journal of Legal Philosophy* 46(2): 161-215.

Caruso, Gregg D. 2021c. "Rejecting Retributivism: Reply to Leo Zaibert." *The Philosopher* 109(4): 118-126.

Caruso, Gregg D., and Derk Pereboom. 2020. "A Non-Punitive Alternative to Punishment." In *Routledge Handbook of the Philosophy and Science of Punishment*, edited by Farah Focquaert, Bruce Waller, and Elizabeth Shaw, 355-365. New York: Routledge.

Corrado, Michael L. 2017. "Punishment and the Burden of Proof." *UNC legal studies research paper*. Available at SSRN: https://ssrn.com/abstract=2997654 or https://doi.org/10.2139/ssrn.2997654.

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

Dingwall, Gavin. 2008. "Deserting Desert? Locating the Present Role of Retributivism in the Sentencing of Adult Offenders." *The Howard Journal of Crime and Justice* 47(4): 400-410.

Husak, Douglas. 2000. "Holistic Retributivism." *California Law Review* 88: 991-1000.

Jeppsson, Sofia. 2020. "Retributivism, Justification and Credence: The Epistemic Argument Revisited." *Neuroethics* 14(2): 177-190.

Kant, Immanuel. 1790 [1991]. *The Metaphysics of Morals*. Translation by Mary J. Gregor. New York: Cambridge University Press.

Kershnar, Stephen. 2000. "A Defense of Retributivism." *International Journal of Applied Philosophy* 14(1): 97-117.

Kershnar, Stephen. 2001. *Desert, Retribution, and Torture*. Lanham, MD: University Press of America.

Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. New York: Oxford University Press.

Mabbott, J.D. 1939. "Punishment." *Mind* 48: 150-167.

Moore, Michael S. 1987. "The Moral Worth of Retribution." In *Punishment and Rehabilitation*, 3rd ed., edited by J.G. Murphy, 94-130. New York: Wadsworth Publishing Company.

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

Moore, Michael S. 1997. *Placing Blame*. New York: Oxford University Press.

Nussbaum, Martha. 2011. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: The Belknap Press of Harvard University Press.

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

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

Pereboom, Derk. 2021. *Wrongdoing and Moral Emotions* New York: Oxford University Press.

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

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

Powers, Madison, and Ruth Faden. 2006. *Social Justice: The Moral Foundations of Public Health and Health Policy*. New York: Oxford University Press.

Quinton, Anthony. 1954. "On Punishment." *Analysis* 14(6): 1933-1942.

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

Robinson, Paul H, and Michael. T. Cahill. 2006. *Law Without Justice: Why Criminal Law Doesn't Give People What They Deserve*. New York: Cambridge University Press.

Sen, Amartya. 1985. *Commodities and Capabilities*. New York: Oxford University Press.

Sen, Amartya. 1999. *Development as Freedom*. New York: Oxford University Press.

Shaw, Elizabeth. 2014. *Free Will, Punishment, and Criminal Responsibility*. Dissertation Thesis. Edinburgh University.

Shaw, Elizabeth. 2021. "The Epistemic Argument Against Retributivism." *Journal of Legal Philosophy* 46(2): 155-160.

Strawson. Galen. 1986. *Freedom and Belief*. New York: Oxford University Press.

Tabandeh, Armin, Paolo Gardoni, and Colleen Murphy. 2017. "A Reliability-Based Capabilities Approach." *Risk Analysis* 38(2): 410-424.

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

Vilhauer, Benjamin. 2009. "Free Will and Reasonable Doubt." *American Philosophical Quarterly* 46(2): 131-140.

Vilhauer, Benjamin. 2012. "Taking Free Will Skepticism Seriously." *The Philosophical Quarterly* 62(249): 833-852.

Vilhauer, Benjamin. 2015. "Free Will and the Asymmetrical Justification of Holding Responsible." *The Philosophical Quarterly* 65(261): 772-289.

Von Hirsch, Andrew. 1976. *Doing Justice: The Choice of Punishment*. Boston: Northeastern University Press.

Von Hirsch, Andrew. 1981. "Desert and Previous Convictions in Sentencing." *Minnesota Law Review* 56(841): 591-634.

Von Hirsch, Andrew. 2007. "The 'Desert' Model for Sentencing: Its Influence, Prospects, and Alternatives." *Social Research* 74(2): 413-434.

Von Hirsch, Andrew. 2017. *Deserving Criminal Sentences: An Overview*. Portland, Oregon: Hart Publishing.

Walen, Alec. 2014. "Retributive Justice." *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/justice-retributive/.

Waller, Bruce. 2011. *Against Moral Responsibility*. Cambridge, MA: MIT Press.

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