Neuroethics
https://doi.org/10.1007/s12152-018-9357-8

CrossMark

Justice without Retribution: An Epistemic Argument against Retributive Criminal Punishment

Gregg D. Caruso

Received: 3 January 2018 / Accepted: 7 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Within the United States, the most prominent justification for criminal punishment is *retributivism*. This retributivist justification for punishment maintains that punishment of a wrongdoer is justified for the reason that she *deserves* something bad to happen to her just because she has knowingly done wrong—this could include pain, deprivation, or death. For the retributivist, it is the *basic desert* attached to the criminal's immoral action alone that provides the justification for punishment. 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. A number of sentencing guidelines in the U.S. have adopted desert as their distributive principle, and it is increasingly given deference in the "purposes" section of state criminal codes, where it can be the guiding principle in the interpretation and application of the code's provisions. Indeed, the American Law Institute recently revised the Model Penal Code so as to set desert as the official dominate principle for sentencing. And courts have identified desert as the guiding principle in a variety of contexts, as with the Supreme Court's enthroning retributivism as the "primary justification for the death penalty." While retributivism provides one of the main sources of justification for punishment within the criminal justice system, 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. In the first section, I explore the retributivist justification of punishment and explain why it is inconsistent with *free will skepticism*. In the second section, I then argue that even if one is not convinced by the arguments for free will skepticism, there remains a strong *epistemic argument* against causing harm on retributivist grounds that undermines both libertarian and compatibilist attempts to justify it. I maintain that this argument provides sufficient reason for rejecting the retributive justification of criminal punishment. I conclude in the third section by briefly sketching my *public health-quarantine model*, a non-retributive alternative for addressing criminal behavior that draws on the public health framework and prioritizes *prevention* and *social justice*. I argue that the model is not only consistent with free will skepticism and the epistemic argument against retributivism, it also provides the most justified, humane, and effective way of dealing with criminal behavior.

Keywords Retributivism · Punishment · Criminal law · Criminal justice · Free will · Moral responsibility · Free will skepticism · Compatibilism · Libertarianism · Desert · Public health · Quarantine · Basic desert · Justice · Social justice · Prison · Retributive justice · Burden of proof

G. D. Caruso (✉)
SUNY Corning, 152 Watauga Avenue, Corning, NY 14830, USA
e-mail: gcaruso@corning-cc.edu

Published online: 15 March 2018

Within the United States, one of the most prominent justifications for legal punishment is *retributivism*. This retributivist justification for punishment maintains that punishment of a wrongdoer is justified for the reason that she *deserves* something bad to happen to her just because she has knowingly done wrong—this could include pain, deprivation, or death. For the retributivist, it is the *basic desert* attached to the criminal's immoral action alone that provides the justification for punishment. 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. A number of sentencing guidelines in the U.S. have adopted desert as their distributive principle [1, 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 dominate principle for sentencing [5, 6]. And courts have identified desert as the guiding principle in a variety of contexts [7–9], as with the Supreme Court's enthroning retributivism as the "primary justification for the death penalty" [10].[1]

While retributivism provides one of the main sources of justification for punishment within the criminal justice system, 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. In first section, I explore the retributivist justification of punishment and explain why it is inconsistent with *free will skepticism*. In second section, I then argue that even if one is not convinced by the arguments for free will skepticism, there remains a strong *epistemic argument* against causing harm on retributivist grounds that undermines both libertarian and compatibilist attempts to justify it. I maintain that this argument provides sufficient reason for rejecting the retributive justification of legal punishment. I conclude in third section by briefly sketching my *public health-quarantine model*, a non-retributive alternative for addressing criminal behavior that draws on the public health framework and prioritizes *prevention* and *social justice* [12–15]. I argue that the model is not only consistent with free will skepticism and the epistemic argument against retributivism, it also provides the most justified, humane, and effective way of dealing with criminal behavior.

[1] See [11, ch.7] for more references and details.

Retributivism and Free Will

To begin, it is important to note that retributive punishment is grounded in the notion of *desert*—the idea that punishing people for certain offenses is morally permissible because such people *deserve* to be punished. As retributivist Mitchell Berman puts it: "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" [16: 269]. Michael S. Moore, another leading retributivist, echoes these sentiments and further highlights the purely backward-looking nature of desert:

> [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 a 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 good effect. [17: 153; see also 18, 19]

Similar accounts of retributivism can be found in Immanuel Kant [20], Stephen Kershnar [21, 22], and Douglas Husak [23]. It's important to note that the concept of *desert* being invoked by retributivists is *basic* in the sense that it is not reducible to consequentialist considerations for punishment or to forward-looking goods such as the safety of society or the moral improvement of the criminal. For the retributivist, since human beings are (sometimes) morally responsible in

the *basic desert* sense [see 24–26], we are justified in giving them their *just deserts* in the form of punishment for moral and legal offenses.

Depending on how retributivists view the relationship between desert and punishment, we can identify three different varieties of the view—*weak*, *moderate*, and *strong* [7–10, 27]. *Weak retributivism* maintains that negative desert (which is what the criminal law is concerned with) is merely necessary but not sufficient for punishment. *Moderate retributivism* maintains that negative desert is necessary and sufficient for punishment but that desert does not mandate punishment—i.e., there may be other goods that outweigh punishing the deserving or giving them their *just deserts* [28]. *Strong retributivism*, on the other hand, maintains that desert is necessary and sufficient for punishment and mandates punishment.[2]

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—e.g., Kant [20], Moore [17–19], Kershnar [21, 22], Husak [23], Berman [16], etc.—and it's my desire to address the dominant view, not a subordinate view held by few. 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" [27: 7]—in fact, some theorists simply define retributivism in a way that excludes *weak retributivism* from consideration altogether.[3] Lastly, the weight the criminal law gives desert and the way retributivism is practically implemented in the law (especially in the U.S.) indicates that the desert of offenders is typically seen as sufficient for punishment. 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 an offender for no purpose other than to see the guilty get what they deserve.[4]

I will also be limiting my investigation to the question of *legal punishment*. Punishment is the intentional imposition of an unpleasant penalty or deprivation for wrongdoing upon a group or individual, typically meted out by an authority. In the case of legal punishment, this authority is the state. More precisely, we can say that legal punishment consists in one person's deliberately harming another on behalf of the state in a way that is intended to constitute a fitting response to some offense and to give expression to the state's disapproval of that offense [29]. Legal punishment is distinct from interpersonal punishment or punishment carried out by an angry mob since it is guided by a system of criminal laws concerned with punishment of individuals who have committed crimes. The question under discussion in this paper, then, is whether *retributive legal punishment* is ever justified?

One reason to think that agents do not *deserve* to suffer for the wrongs they have done in the purely backward-looking sense required for retributivism is that they are not free and morally responsible in the sense required. *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—see, e.g., Derk Pereboom [24, 25], Galen Strawson [30, 31], Saul Smilansky [32], Neil Levy [33], Bruce Waller [34, 35], and myself [36].

[2] Immanuel Kant, for example, famous maintained 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 take 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. 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" [20: Part II: 6].

[3] David Boonin, for example, defines retributivism as follows: "It claims 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" [29: 85 (emphasis added)].

[4] I will also henceforth drop the distinction between *moderate* and *strong* retributivism. Strong retributivists defend two distinct claims: that we are justified in punishing offenders and that we have a duty to do so. Moderate retributivists seek only to defend the first claim. Since I am simply concerned with the question, "Does the desert of offenders sometimes provide sufficient grounds for punishment?" we can set this distinction aside here.

Most argue that while determinism is incompatible with free will and basic desert moral responsibility, so too is *indeterminism*, especially the variety posited by quantum mechanics [e.g., 24, 25, 36]. Others argue that regardless of the causal structure of the universe, we lack free will and moral responsibility because free will is incompatible with the pervasiveness of *luck* [33]. Others (still) argue that free will and ultimate moral responsibility are incoherent concepts, since to be free in the sense required for ultimate moral responsibility we would have to be *causa sui* (or "cause of oneself) and this is impossible [30, 31].

What all these arguments 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 [25, 37, 38]. Nor is it to deny that there may be good pragmatic reasons to maintain certain systems of punishment and reward. 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. Hence, free will skepticism presents a challenge to retributivism since it does away with the idea of basic desert. If agents do not deserve blame just because they have knowingly done wrong, neither do they deserve punishment just because they have knowingly done wrong [25: 157].

While I have elsewhere argued that our best philosophical and scientific accounts of the world support free will skepticism [15, 36, 39], the *epistemic argument* I develop below remains independent of this claim. The argument maintains that even if one is not convinced by the arguments for free will skepticism 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 retributivists 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 severe. 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 that the skeptical 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.

Epistemic Argument

I will now argue that the retributive justification of legal punishment fails to meet the burden of proof required to justify the harms caused. My argument for this conclusion runs as follows:

(1) Legal punishment 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 [40].

(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.

This argument builds on previous work done by Double [41], Vilhauer [42], Corrado [43], and Pereboom [24, 25, 40], and it maintains that no extant account of basic desert moral responsibility has the evidentiary support needed to justify retributive legal punishment.[5]

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. It maintains that if it significantly probable that one's justification for harming another is unsound, then, *prima facia*, that behavior is seriously wrong. Support for this premise 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. [43: 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. Lets 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 in intentional, like when a police officer decides to use deadly force. The use of deadly force, we demand, may be justified only under conditions of extreme necessity, when all lesser means have failed or cannot reasonably be employed. And while we may acknowledge that the level of justification required may vary depending on the severity of harm involved, we typically demand that one must have good justification for intentionally harming another.

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 [45] as well as Corrado [43: 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:

> …in a purely philosophical contest where nothing of a practical nature hangs on the outcome it is the evidential burden of proof that is required, and the standard of proof must be "by a preponderance of the evidence": whoever simply has the better evidence must win. On the other hand, if something practical does depend on the outcome of the philosophical debate, then what would matter is the prudential burden. The costs on either side would determine the allocation of the burden and the standard by which satisfaction of the burden is to be measured. [43: 3]

I contend that given the practical important of moral responsibility to legal punishment, and given the gravity of harm caused by legal punishment (to the individuals

[5] I should note that Double limits his argument to libertarianism (see below) and Vilhaeur stops short of extending his concerns to legal punishment. While Corrado's argument is similar to the one I develop here, we both arrived at our arguments independently. I first sketched the basic structure of the epistemic argument back in 2015 when working on my book manuscript *Unjust Deserts: Free Will, Moral Responsibility, and Legal Punishment* (still unpublished) and briefly referenced it in Caruso [44]. Pereboom [24, 25, 40] also runs a similar argument but does not develop it in detail.

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.

When premise (1) is combined with (2), which is simply a statement of the retributivist justification for legal punishment as summarized in the previous section, we get the requirement that retributivists must justify their core assumption—i.e., that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done. While this demand for justification is reasonable given the strength of (1), many retributivists simply deny or ignore it. For instance, Larry Alexander, Kimberly Kessler Ferzan, and Stephen Morse—in their book *Crime and Culpability: A Theory of Criminal Law*—maintain that "[w]e need take no stand on the freewill-determinism issue" [27: 15].[6] They go on to explain:

> Two of is—Ferzan and Morse—are persuaded by the arguments for compatibilism. One of us—Alexander—is not. His view is that compatibilism provides only a hollow form of moral responsibility, not the full-blooded form that our reactive attitudes assume. In particular, it seems unresponsive to the worry that what appears to an actor to be a reason, or a reason with a particular positive or negative weight, seems to be beyond the actor's proximate control. One the other hand, he also believes that libertarianism cannot deliver a form of moral responsibility worth wanting because, just like determinism, its foil, libertarianism takes control out of the agent's hands and relinquishes it to chance—or else just makes it utterly mysterious. [27: 15]

How, then, does Alexander justify doing harm on retributive grounds when "neither determinism nor indeterminism can provide a satisfactory account of moral responsibility, and together they appear to exhaust the possibilities" [27: 15]? We are provided with the following, rather unsatisfying, answer:

> Alexander believes, as a metaphilosophical position, that the freewill-determinism puzzle is one of those antinomies of thought that we are incapable of resolving, along with the mind-body and infinity puzzles. For him, the freewill-determinism puzzle will always dog practices of holding people morally responsible, practices that we nevertheless cannot dispensing with. *Because we cannot dispense with such practices, a retributivist regarding criminal punishment need not resolve or even take sides on the freewill issue*. [27: 15 (italics added)]

Alexander's position seems to be that while we will never be able to comprehend the bases of moral responsibility [see 27: 15n22], we should nevertheless continue causing harm on retributivist grounds since "we cannot dispense with such practices."

There are, however, at least two major problems with this reply. First, why think that we can never dispense with our retributive practices when it comes to the criminal law? If Alexander were limiting his comments to our interpersonal reactive attitudes, arguing (à la Strawson [46]) that they were indispensible or centrally important, that would be one thing. But in the comment above and the book from which it is drawn, Alexander is discussing legal punishment. I see no reason to think that retributivism is indispensible to the criminal law. The fact that several countries have moved away from retributive models of legal punishment (to greater or lesser extent) is evidence enough that this is false. It is also important to distinguish between *narrow-profile* and *wide-profile* reactive attitudes [25, 47]. Narrow-profile attitudes are local or immediate emotional reactions to situations, whereas wide-profile responses are not immediate and involve rational reflection. While we may be unable to appreciably reduce narrow-profile retributive reactions in some cases, it is open for us to diminish or even eliminate retributive reactions in wide-profile cases. If I am hurt in an intimate personal relationship, for example, it may be beyond my ability to resist feeling resentment or anger. But when it comes to the law, which involves wide-profile rational reflection, we can indeed disavow retributivism in the sense of rejecting any force it may be assumed to have in *justifying* a harmful response to wrongdoing.

[6] In the book, Alexander, Ferzan, and Morse set out to explain what the criminal law would look like if structured by moderate retributivism: "What we intend to do in this book is to explore what the doctrines of the criminal law would look like if they were structured (primarily) by the concern that criminal defendants receive the punishment they deserve, and particularly that they receive no more punishment than they deserve…In our view, it is the defendant's decision to violate society's norms regarding the proper concern due to the interests of others than establishes the negative desert that in turn can both justify and limit the imposition of punishment" [27: 6–7].

Second, to continue doing harm on retributivist grounds even though one acknowledges that the philosophical foundations of the view can never be comprehended or justified, violates the basic precept, both widespread and intuitive, that one should restrain from doing harm unless otherwise justified. Note that Alexander is *not* saying that despite our inability to justify, in some foundational way, basic desert moral responsibility, we should consider libertarianism or compatibilism as providing *justification enough*. Instead, he explicitly states, "compatibilism provides only a hollow form of moral responsibility" and "libertarianism cannot deliver a form of moral responsibility worth wanting." Since this is a conclusion shared by many free will skeptics, it's unclear why he thinks we should continue holding agents morally responsible in the basic desert sense, let alone subject them to severe and often painful forms of legal punishment. It in no way helps to claim that the problem of free will is an antinomy. One does not get off the justificatory hook with regard to legal punishment by simply throwing their hands up in the air and saying, "since there is no satisfying solution to the problem of free will we should continue assuming agents are morally responsible and hence justly deserve to suffer for the wrongs they have done." Rather, one needs to provide a positive and compelling reason *for* believing that agents are free and morally responsible in the sense required.

Morse and Ferzan at least claim they are "persuaded by the arguments for compatibilism." But how persuaded? In answering this, I will focus on Morse's take on compatibilism since he spells it out at great length elsewhere [48–51]. He begins by noting that:

> The criminal law is a thoroughly folk-psychological enterprise. Doctrine and practice implicitly assume that human beings are agents, creatures who act intentionally for reasons, who can be guided by reasons, and who in adulthood are capable of sufficient rationality to ground full responsibility unless an excusing condition obtains. We all take this "standard picture" for granted because it is the foundation not just of law but of interpersonal relations generally, including how we explain ourselves to others and to ourselves. [49: 40]

He goes on to argue that the standard picture is thoroughly compatibilist in nature. He acknowledges that "metaphysical assumptions matter" [49: 44], especially when it comes to the criteria for moral responsibility, but adds: "The question is whether one must resolve or even defend one's metaphysical and other philosophical foundations in these fraught areas. I think not" [49: 44]. According to Morse, "when philosophy is foundational and practically important, one's position must be acknowledged but need not be defended or, *a fortiori*, resolved" [49: 45].

When it comes to the free will debate, Morse maintains that "[t]here will always be good arguments for and against the various positions" [49: 45]. With no ability to declare a winner, he asks himself: "What is a poor, country lawyer-scholar to do in such circumstances when trying to make normative arguments about doctrine, practice, and policy?" [49: 45]. His answer is to:

> …start with a normative position that is attractive at the non-metaphysical level of applied ethical, moral, political, and legal theory. If this position is consistent with a reasonable metaphysics that does not conflict with relatively uncontroversial, or at least plausible, empirical accounts about the world and with other reasonable philosophical theories, then one can proceed without defending the metaphysics, the empirics, and other philosophical positions. [49: 46]

According to Morse, compatibilism provides just such a reasonable metaphysics. And while incompatibilists are likely to point to all the arguments against compatibilism that suggest that the normative position adopted is unjustified, Morse maintains that "a sophisticated metaphysician who adheres to the chosen metaphysics would have answers, and there would be no decisive arguments to refute the sophisticate" [49: 46]. It is this lack of decisiveness that Morse points to as key. He also maintains that any position that wants to deny the "standard picture" should carry the burden of proof, since "[a]ny position that violates common sense should meet the most demanding burden of persuasion" [49: 46]. For these reasons, he concludes: "I am compatibilist, a perfectly plausible metaphysics, and will continue to believe that robust responsibility is possible until an incontrovertible argument that all would accept requires me to jettison this view" [49: 49].

While Morse's reasoning is quite sophisticated I contend that it gets things exactly backwards. First, free will skepticism also offers a reasonable metaphysics, one

that is consistent with our best philosophical and scientific accounts of the world. While Morse might find it is unattractive at the level of applied moral, political, and legal theory [51], Pereboom, Waller, and I have elsewhere argued that adopting the skeptical perspective actually has distinct advantages in these areas [12, 13, 15, 37, 39, 52; see also 24, 25, 34, 35]. Second, I disagree with Morse on who carries the burden of proof. He claims the burden falls on those who want to reject basic desert moral responsibility (what he calls "robust responsibility). I, on the other hand, maintain that if one is going to cause harm on the assumption that compatibilism is true and agents are morally responsible in the sense required for retributive justice, then the burden is on them to establish beyond a reasonable doubt that this is the case. While Morse is correct that in *certain circumstances* the burden falls on those who advance philosophical views contrary to common sense, things change dramatically when significant harm is involved. Perhaps it's acceptable to adopt a reasonable metaphysics with regard to, say, realism about the external world, but this is significantly different than adopting compatibilism about moral responsibility since it is not used to justify retributive harm.

My proposal is that we adopt a *precautionary principle* when it comes to unresolved metaphysical/philosophical positions that are likely to cause severe harm. The principle would be analogous to the ones adopted by the United Nations General Assembly, various legally binding international treaties (e.g., the Montreal Protocol, Rio Declaration, and Kyoto Protocol), and the law of the European Union. The precautionary principle has been used by policy makers to justify discretionary decisions in situations where there is the possibility of harm from a decision, course of action, or policy, and extensive scientific knowledge on the matter is lacking. The principle implies that there is a social responsibility to protect the public from exposure to harm, when scientific investigation has found a plausible risk.[7] The principle has been adopted in numerous contexts, including in response to climate change skeptics who argue that we should do nothing to cap Co2 and other greenhouse gas emissions in the face of (what they consider to be) scientific uncertainty about man-made climate change. The 1992 United Nations Framework Convention on Climate Change, for instance, states: "When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically."

In the current context, I propose that we employ a precautionary principle that constrains which "reasonable metaphysics" one adopts in the face of philosophical uncertainty—or, at least, constrains the practical implementation of that metaphysics in the arena of criminal law. If, as Morse thinks, the free will debate is currently unresolvable (something I disagree with but will not take issue with here), then on precautionary grounds we should prohibit causing retributive harm in the case of legal punishment until and unless the high epistemic standard of proof outlined earlier is met. Just like tobacco executives who challenged the link between smoking and cancer by appealing to scientific uncertainty, desertists who justify retributive harm on the grounds that no "incontrovertible argument that all would accept" has yet been presented against the notions of desert and moral responsibility are violating a pretty reasonable precautionary principle. The principle demands that in the context of uncertainty, precautionary measures should be taken to protect individuals from unjustified harm.

But what about those philosophers who are more confident about the existence of free will and basic desert moral responsibility? Both Morse and Alexander appear willing to accept premise (4) of the epistemic argument, which states that the philosophical arguments for libertarianism and compatibilism are subject to powerful and unresolved objections. They instead reject the demand for a high level of justification for the claim that agents are morally responsible and hence justly deserve to suffer for the wrongs they have done. Yet philosophers who defend libertarian and compatibilist accounts of free will are more likely to challenge premise (4). In defending premise (4) against these accounts, I will take as my starting point Bruce Waller's powerful insight—spelled out and defended at great length in his recent book *The Stubborn System of Moral Responsibility* [35]—that *belief* in basic desert moral responsibility is stronger than the philosophical arguments presented in their favor. More specifically, I will argue that the philosophical arguments presented in favor of basic desert moral responsibility, which are needed to justify retributive harm, are either scientifically implausible (as in the case of agent causation), empirically unwarranted (as in the case of event causal libertarianism), beg the question

[7] See *Wikipedia* entry for "precautionary principle": https://en.wikipedia.org/wiki/Precautionary_principle

(as in the case of Strawson and other forms of compatibilism), or end up "changing the subjection" (as in the case of Dennett and others).

Let me being with a rather obvious case of a philosopher who believes more strongly in moral responsibility than the philosophical arguments he presents in its favor, that of Peter van Inwagen. After championed the *consequence argument* in favor of incompatibilism,[8] van Inwagen proceeds to argue that we *must* reject determinism even though it means free will "remains a mystery" [53, 54]. He acknowledges that libertarian free will is difficult to make sense of, yet he also claims that to "deny the free-will thesis is to deny the existence of moral responsibility, which would be absurd" [53: 223]. According to van Inwagen, "free will undeniably exists and there is a strong and unanswered prima facie case for its impossibility" [54: 1–2]. It is the *absurdity* of denying moral responsibility that leads van Inwagen to favor libertarianism (and embrace a mystery) over its incompatibilist alternative of free will skepticism. But he doesn't stop there. He continues on to say that, *if* science were one day able to present us with compelling reasons for believing in determinism, "[t]hen, and only then, I think, should we become compatibilists" [53: 223]. Essentially, after several decades of heroically defending the consequence argument, van Inwagen would be willing to chuck it out the window (if need be) to preserve moral responsibility. Such a defense of moral responsibility (if one can call it that) is a far cry from the level of justification needed to license the kind of harm caused by retributive legal punishment—which is all I am concerned with here.

Additional evidence of the kind of stubbornness Waller has in mind can be found among agent-causal libertarians—such as C. A. Campbell [55], Richard Taylor [56], and Roderick Chisholm [57]—who are willing to embrace mysterious and "god-like" powers and abilities to preserve moral responsibility. Chisholm, for example, famously argued: "If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved" [57: 32]. While Chisholm appears to gleefully embrace such miraculous powers, Taylor is at least more willing to acknowledge the embarrassment of such a move. He says, for instance, of his own theory:

> One could hardly affirm such a theory of agency with complete comfort, however, and not wholly without embarrassment, for the conception of agents and their powers which is involved in it is strand indeed, if not positively mysterious. In fact, one can hardly be blamed here for simply denying our data outright, rather than embrace this theory to which they do most certainly point. [56: 53]

This is far cry form the epistemic standard needed to justify punishing someone on the grounds that they posses libertarian free will and therefore deserve it. And as Waller so eloquently points out: "When contemporary philosophers are willing to posit miracles in order to save moral responsibility, the philosophical belief in moral responsibility obviously runs deep and strong" [35: 3].

Naturalistically minded event-causal libertarians have the advantage of avoiding miraculous *sui generis* kinds of causal powers, but when it comes to providing the epistemic justification needed to ground retributive punishment they too fall far short. Consider, for instance, the influential accounts of Robert Kane [58] and Mark Balaguer [59]. Neither philosopher claims to have proven libertarian free will exists or that the necessary empirical requirements posited on their respective accounts actually obtain. Rather, both settle for the much weaker claim that their theories are *consistent* with our best scientific theories and have *not yet* been ruled out. The *mere possibility* that one of these accounts may be true, however, is simply not enough to provide the epistemic justification needed to ground retributive harm. This becomes especially clear when one considers what *actually* needs to be the case for, say, Kane's account to be true. As Manuel Vargas writes:

> [N]ot only do agent mental processes have to turn out to be indeterministic, but they must also be indeterministic in a very particular way. If multiple mutually exclusive aims did not cause the brain to go into a chaotic state the theory would be disproved. If it turned out that neurological systems weren't sensitive to quantum indeterminacies the theory would be disproved. If it turned

[8] Although there are many different formulations of the consequence argument, van Inwagen summarizes the basic idea as follows: "If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequence of these things (including our present acts) are not up to us" [53: 16].

> out that neurological systems were sensitive to quantum indeterminacies, but not sufficiently sensitive to amplify quantum indeterminacies in a way that affects the outcomes of choice, this too would disprove the theory. These are not marginal or insubstantial bets about what brain science will reveal to us. [60, 143]

I should add, if it turned out that neurological systems were sufficiently sensitive to amplify quantum indeterminacies in a way that affected the outcomes of choice, but such indeterminacies did not arise at the right *temporal* moment required, the theory would also be disproven. To justify retributive practices, including in some instances excessively punitive practices, on the off chance that all these empirical conditions will be met is akin to moral malpractice.

The fact that libertarian accounts of free will fail to provide the epistemic justification needed to ground retributivism, yet continue to hold agents morally responsible in the basic desert sense, is why Richard Double famously charges them with "hard-heartedness" [41]. Double's argument can be summarized as follows. Any thinker who holds the following three theses is hard-hearted (i.e., morally unsympathetic, not morally conscientious): (1) We may hold persons morally responsible only if they make libertarian choices; (2) we should hold persons morally responsible; and (3) there is scant epistemic justification that persons make libertarian choices. Almost all libertarians endorse (1) and (2), and most would acknowledge (3) if pushed on it. To the extent, then, that (3) is true (and I believe it is), it would be hard-hearted for libertarians to hold persons morally responsible. The reasoning here is analogical. Just as it would be hard-hearted to punish subjects for actions if we lacked strong evidence that those subjects did those actions, it would likewise be hard-hearted to blame subjects (in the backward-looking sense) for actions if we lacked strong evidence that those subjects did those actions from free will.

While Double restricts his charge of hard-heartedness to libertarians, I think a similar argument can be made against compatibilists who want to continue harming individuals on retributivist grounds. In the case of compatibilism, however, the lack of epistemic justification comes at a different point than that of libertarianism. The epistemic challenge facing libertarianism is to justify the claim that we *actually* posses the powers and abilities posited by such accounts. There is no equivalent debate with regard to compatibilism. All parties agree, including skeptics, that we have the abilities discussed by most leading compatibilist accounts—e.g., reasons-responsiveness, voluntariness, the capacity to act in accordance with moral reasons or one's Deep Self, etc. The question instead is whether such abilities are *enough* to justify basic desert moral responsibility and along with it retributive harm in the case of legal punishment. It is here that scant epistemic justification is generally provided.

While compatibilists reject miracles and propose accounts of moral responsibility consistent with both naturalism and determinism, they seldom provide justification for the moral responsibility system itself. In lieu of justifying the moral responsibility *system*, compatibilists typically take the system as given and instead focus on what attitudes, judgments, and treatments are justified from *within* the system. P. F. Strawson [46] is a good example of this. His defense of the reactive attitudes takes our normal moral responsibility practices as given and proceeds from there to articulate special circumstances when it is acceptable *not* to hold someone morally responsible or to excuse them—e.g., when they are profoundly impaired by delusion or lack any moral capacity, either temporarily or permanently. In these special circumstances, Strawson claims it is acceptable to adopt the *objective attitude* toward them. But according to Strawson and his followers, the denial of *all* moral responsibility is unacceptable, self-defeating, and/or impossible, since to permanently excuse everyone would entail that "nobody knows what he's doing or that everybody's behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense…" [46: 74].

The problem with this defense of moral responsibility, however, is that it takes for granted the very thing in need of justification. To quote Waller once again:

> [I]f we start from the assumption of the moral responsibility system (assumptions that are so common and deep that they are difficult to escape), then the denial of moral responsibility is absurd and self-defeating. But the universal denial of moral responsibility does *not* start from the assumption that under normal circumstances we are morally responsible, and it does *not* proceed from that starting point to enlarge and extend the range of excuses to cover everyone (so that *everyone* is profoundly flawed). That is indeed a path to

> absurdity. Rather, those who reject moral responsibility reject the basic system which starts from the assumption that all minimally competent persons (all who reach the plateau level) are morally responsible. For those who deny moral responsibility, it is never fair to treat anyone as morally responsible, no matter how reasonable, competent, self-efficacious, strong-willed, and clear-sighted that person may be. [35: 103]

Since skeptics who globally challenge the moral responsibility system—e.g., Waller [34, 35], Pereboom [24, 25], Levy [33], G. Strawson [30, 31], and myself [36]—do not accept the rules of that system, it is question begging to assume our ordinary moral responsibility practices are justified without refuting the arguments for global skepticism.

Furthermore, even if Strawson were correct about the necessity of holding each other responsible in our interpersonal exchanges, it's another thing altogether to think retributive legal punishment is justified. Perhaps one could justify adopting an evidential burden of proof by limiting one's compatibilism to our interpersonal reactive attitudes where the level of harm involved is significantly lower than legal punishment (though still not inconsequential). But in the case of legal punishment, retributivists who adopt compatibilism need to meet a higher burden of proof. It also does not help if one decides to shift the focus away from Strawson's reactive attitudes and toward an account of what those attitudes might rightfully be reacting to—namely some capacity or feature of the agent. While contemporary compatibilists have conceived of this capacity in a number of different ways—including the capacity to act in accordance with one's Deep Self, the capacity for rational review of actions, the capacity to act in accordance with moral reasons, and the capacity for reasons-responsiveness—the central question remains: Are these capacities (alone or combined) *enough* to ground basic desert moral responsibility? Given that extant compatibilist accounts still face powerful and unresolved objections—such as the manipulation argument [61–63], Pereboom's four-case argument [24, 25], van Inwagen's consequence argument [53], Galen Strawson's basic argument [30, 31], Fischer's no-forking-paths argument [64], Levy's luck pincer [33], etc.—they fail to meet the prudential burden of proof. For a retributivist to assume that compatibilism is true and proven beyond a reasonable doubt is to beg the question against incompatibilists and permit unjustified harm.

While some (or most) forms of compatibilism beg the question, others simply change the subject. Daniel Dennett is a good example of the latter since the kind of free will he argues for does not attempt to justify retributivism or what I have called basic desert moral responsibility. In a revealing exchange with Tom Clark and Bruce Waller—on Waller's book *Against Moral Responsibility* [34]—Dennett makes this point clear. After explaining how Waller rejects retributive punishment and basic desert moral responsibility, Dennett writes:

> [L]et me say right away that I agree with Waller's main conclusion in one important sense: *that* kind of absolutistic moral responsibility—insisting as it does on what I have called guilt-in-the-eyes-of-God—is incompatible with naturalism and has got to go. [65]

Rejecting what he considers to be an untenable retributivist system of legal punishment and the strong type of basic desert moral responsibility that it relies upon, Dennett proposes "a consequentialist defense of just deserts" [65] and punishment. While many compatibilists want to secure a justification for retributive blame and punishment, Dennett does not seem to be among them. Instead, he espouses a non-retributivist, consequentialist concept of moral responsibility and punishment. What is of paramount importance for Dennett is that we find a naturalistic way to justify punishment so as to maintain a "secure and civil society" [65]. The justification Dennett provides for punishment, however, is decidedly different than what many compatibilists would endorse since he acknowledges that we must abandon retributivism and basic desert moral responsibility and replace it with an "ultimately consequentialist" [65] conception of punishment and reward. Given that most free will skeptics would agree, Dennett's account fails to preserve what is of central philosophical and practical importance—the claim that people deserve to be praised and blamed, rewarded and punished, on strictly retributivist and non-consequentialist grounds [see 26]. In fact, Dennett admittedly does not attempt to preserve this sort of moral responsibility and cheerfully wishes it "good riddance" [65].

One important difference, however, between Dennett's consequentialist justification of punishment and that of free will skeptics is that Dennett prefers to retain the notion of *just deserts*. I contend that this is inconsistent with Dennett's reformed consequentialist account of moral responsibility. As Tom Clark notes:

> Whether as consequentialists we should still talk of just deserts is debatable, given the strong deontological, retributive connotations…What you're advocating is the *practical necessity* of punishment, not its intrinsic goodness, but "just deserts" strongly implies that the offender's suffering is intrinsically good, which you don't think is the case. So I think we should drop talk of just deserts so we don't mislead people about what we believe are defensible justifications for punishment. [66]

Given the canonical understanding of *just deserts* and how it is used to justify various retributive attitudes, judgments, and treatments, Dennett's use of the term lends itself to easy confusion and gives the mistaken impression that he is setting out to preserve something that he is not. Rather than defending or justifying just deserts, Dennett ends up changing the subject. His brand of compatibilism does not justify retributive harm—it doesn't even attempt to. His justification for punishment, being consequentialist in nature, is completely consistent with the skeptic's rejection of free will and just deserts.

In this section I've argued that the burden of proof is on the retributivist to justify legal punishment and given the severity of harm involved the highest epistemic standard should be adopted. Drawing from the criminal law, I argued that the proper place to set the epistemic bar is at the prudential standard of beyond a reasonable doubt. I further argued that all extant accounts of libertarianism and compatibilism fail to satisfy this burden of proof since they are either scientifically implausible (as in the case of agent causation), empirically unwarranted (as in the case of event causal libertarianism), question begging (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). Hence, the epistemic argument provides sufficient reason for rejecting retributive legal punishment.

The Public Health-Quarantine Model

If retributive punishment is unjustified, where does that leave us with regard to criminal behavior? In this final section I will suggest that the *public health-quarantine model* provides an ethically defensible and practically workable alternative to retributivism. Given the limited space remaining, I can do no more that sketch the model. For a fuller defense as well as replies to objections, see Caruso [12–15] and Pereboom [24, 25, 40, 67].[9]

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 [24, 25, 67]. 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 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 [25: 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 [15]. 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 [15, 40].

Second, the quarantine model places several constraints on the treatment of criminals [15, 24, 25]. First, as less dangerous diseases justify only preventative measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate

[9] Material in this section has been drawn from Caruso [12–15].

restraints [25: 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 [25: 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 [25: 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 [12–15]. 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* [12–14]. 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 [13].

In my recent *Public Health and Safety: The Social Determinants of Health and Criminal Behavior* [13], 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 [12–14]. 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 [13].

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., Amartya Sen and Martha Nussbaum, see 68–71]. 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* [13, 71]. Following Powers and Faden [71], 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" [71: 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 real opportunity to promote or achieve functionings and "to lead the kind of life he or she has reason to value" [69: 87].

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 [71]. 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 legal 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.

Conclusion

In this paper I have argued that one of the most prominent justifications for legal punishment, retributivism, fails to justify the harms it causes and is therefore seriously wrong. After exploring the retributive justification of legal punishment and explaining why it is inconsistent with free will skepticism (in first section), I argued that even if one is not convinced by the arguments for free will skepticism there remains a strong epistemic argument against causing harm on retributivist grounds that undermines both libertarian and compatibilist attempts to justify it (in second section). I argued that this argument provides sufficient reason for rejecting retributive legal punishment. I then concluded (in third section) by briefly sketching my public health-quarantine model, a non-retributive alternative for addressing criminal behavior that draws on the public health framework and prioritizes prevention and social justice. The public health-quarantine model, I contend, offers a non-retributive way forward and a suitable conception of justice without retribution. The burden of proof is on the retributivist to explain why this alternative is unacceptable and, more importantly, how they overcome the epistemic argument given that reasonable doubt still remains concerning the existence of free will and basic desert moral responsibility.

References

1. 204 Pa. Code sect. 303.11 (2005).
2. Tonry, Michael. 2004. U.S. sentencing systems fragmenting. In Panel reform in overcrowded times, ed. Michael Tonry, 21–28. New York: Oxford University Press.
3. Cal. Penal Code Sect. 1170(a)(1) (West 1985).
4. 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. Caruso, Gregg D. 2017. The American law institute revises the model penal code. *Pyschology Today*, May 26: https://www.psychologytoday.com/blog/unjust-deserts/201705/the-american-law-institute-revises-the-model-penal-code
7. U.S. cases Spaziano v. Florida, 468 U.S. 447, 462 (1984).
8. Gregg v. Georgia, 428 U.S. 153, 183–184 (1976).
9. Cotton, Michele. 2000. Back with a vengeance: The resilience of retribution as an articulated purpose of criminal punishment. *American Criminal Law Review* 37 (4): 1313–1357.
10. *Spaziano v. Florida*, 468 U.S. at 461.
11. Robinson, Paul H. 2008. *Distributive principles of criminal law: Who should be punished how much*. New York: Oxford University Press.
12. Caruso, Gregg D. 2016. Free will skepticism and criminal behavior: A public health-quarantine model. *Southwest Philosophy Review* 32 (1): 25–48.
13. Caruso, Gregg D. 2017. *Public health and safety: The social determinants of health and criminal behavior*. UK: ResearchLinks Books.
14. Caruso, Gregg D. Forthcoming. The public health-quarantine model. In *Oxford Handbook of Moral Responsibility*, eds. Dana Nelkin and Derk Pereboom. New York: Oxford University Press.
15. Pereboom, Derk, and Gregg D. Caruso. 2018. Hard-incompatibilist existentialism: Neuroscience, punishment, and meaning in life. In *Neuroexistentialism: Meaning, morals, and purpose in the age of neuroscience*, ed. Gregg D. Caruso and Owen Flanagan, 193–222. New York: Oxford University Press.
16. Berman, Mitchell. 2008. Punishment and justification. *Ethics* 18: 258–290.
17. Moore, Michael S. 2010. *Placing blame*. New York: Oxford University Press.
18. Moore, Michael S. 1987. The moral worth of retribution. In Punishment and rehabilitation, third edition, ed. Jeffrie G. Murphy, 94–130. Wadsworth Publishing Company.
19. Moore, Michael S. 1993. *Act and crime: The philosophy of action and its implications for criminal law*. New York: Oxford University Press.
20. Kant, Immanuel. 1790. *The metaphysics of morals*. Hastie: Translated by W.
21. Kershnar, Stephen. 2000. A defense of retributivism. *International Journal of Applied Philosophy* 14 (1): 97–117.
22. Kershnar, Stephen. 2001. *Desert, retribution, and torture*. Lanham, MD: University Press of America.
23. Husak, Douglas. 2000. Holistic retributivism. *California Law Review* 88: 991–1000.
24. Pereboom, D. 2001. *Living without free will*. New York: Cambridge University Press.
25. Pereboom, D. 2014. *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
26. Caruso, Gregg D., and Stephen G. Morris. 2017. Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance. *Erkenntnis* 82: 837–855.
27. Alexander, Larry, Kimberly Kessler Ferzan, with Stephen Morse. 2009. Crime and culpability: A theory of criminal law. New York: Cambridge University Press.
28. 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.
29. Boonin, David. 2008. *The problem of punishment*. New York: Cambridge University Press.
30. Strawson, Galen. 1986. *Freedom and belief*. Oxford: Oxford University Press.
31. Strawson, Galen. 1994. The impossibility of moral responsibility. *Philosophical Studies* 75 (1): 5–24.
32. Smilansky, Saul. 2000. *Free will and illusion*. New York: Oxford University Press.

33. Levy, Neil. 2011. *Hard luck: How luck undermines free will and moral responsibility*. New York: Oxford University Press.
34. Waller, Bruce. 2011. *Against moral responsibility*. Cambridge, MA: MIT Press.
35. Waller, Bruce. 2015. *The stubborn system of moral responsibility*. Cambridge, MA: MIT Press.
36. Caruso, Gregg D. 2012. *Free will and consciousness: A determinist account of the illusion of free will*. Lanham, MD: Lexington Books.
37. Caruso, Gregg D. 2017. Free will skepticism and the question of creativity: Creativity, desert, and self-creation. *Ergo* 3 (23): 23–39.
38. Caruso, Gregg D. 2018. Skepticism about moral responsibility. *Stanford Encyclopedia of Philosophy.*
39. Caruso, Gregg D. 2013. Introduction: Exploring the illusion of free will and moral responsibility. In *Exploring the illusion of free will and moral responsibility*, ed. Gregg D. Caruso, 1–16. Lanham, MD: Lexington Books.
40. Derk, Pereboom. 2017. A defense of free will skepticism: Replies to commentaries by victor Tadros, Saul Smilansky, Michael McKenna, and Alfred R. Mele on *Free will, agency, and meaning in life*. *Criminal Law and Philosophy* 11 (3): 617–636.
41. Double, Richard. 2002. The moral hardness of libertarianism. *Philo* 5: 226–234.
42. Vilhauer, Ben. 2015. Free will and the asymmetrical justifiability of holding morally responsible. *The Philosophical Quarterly* 65 (261): 772–789.
43. Corrado, Michael Louis, Punishment and the Burden of Proof (March 3, 2017). UNC legal studies research paper. Available at SSRN: https://ssrn.com/abstract=2997654 or https://doi.org/10.2139/ssrn.2997654
44. Caruso, Gregg D. 2016. Arrested development. *RSA Journal* 2: 42–46.
45. Pigliucci, Massimo, and Maartin Boudry. 2014. Prove it! The burden of proof game in science vs. pseudoscience disputes. *Phiosophia* 42: 487–502.
46. Strawson, P. F. 1982. Freedom and resentment. In *Free will*, ed. G. Watson, 59–80. New York: Oxford University Press. (Originally punished in *Proceedings of the British Academy* 36 [1962]).
47. Nichols, Shaun. 2007. After compatibilism: A naturalistic defense of the reactive attitudes. *Philosophical Perspectives* 21: 405–428.
48. Morse, Stephen J. 2013. Common criminal law compatibilism. In *In Neuroscience and legal responsibility, ed*. Nicole A. Vincent. New York: Oxford University press.
49. Morse, Stephen J. 2015. Criminal law and common sense: An essay on the perils and promise of neuroscience. *Marquette Law Review* 99: 39–72.
50. Morse, Stephen J. 2015. Neuroscience, free will, and criminal responsibility. In *Free will and the brain: Neuroscientific, philosophical, and legal perspectives, ed. Walter Glannon*. Cambridge, MA: Cambridge University Press.
51. Morse, Stephen J. 2018. The neuroscientific non-challenge to meaning, morals, and purpose. In *Neuroexistentialism: Meaning, morals, and purpose in the age of neuroscience*, ed. Gregg D. Caruso and Owen Flanagan, 333–358. New York: Oxford University Press.
52. Caruso, Gregg D. Forthcoming. Free will skepticism and its implications: An argument for optimism. In *Free will skepticism in Law and Society*, eds. Elizabeth Shaw, Derk Pereboom, and Gregg D. Caruso. New York: Cambridge University Press.
53. Van Inwagen, Peter. 1983. *An essay on free will*. New York: Clarendon Press.
54. Van Inwagen, Peter. 2000. Free will remains a mystery. *Philosophical Perspectives* 14: 1–19.
55. Campbell, C. A. 1957. On selfhood and godhood. London: George Allen & Unwin.
56. Taylor, Richard. 1963. *Metaphysics*. Englewood Cliffs, NJ: Prentice-Hall.
57. Chisholm, Robert. 1982. Human freedom and the self. In *Free will*, ed. Gary Watson, 24–35. New York: Oxford University Press.
58. Kane, Robert. 1996. *The significance of free will*. New York: Oxford University Press.
59. Balaguer, Mark. 2009. *Free will as an open scientific problem*. Cambridge, MA: MIT Press.
60. Vargas, Manuel. 2007. Revisionism. In *Four views on free will, ed. John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas*, 126–165. Oxford: Blackwell.
61. Mele, Al. 1995. *Autonomous agents*. New York: Oxford University Press.
62. Mele, Al. 2006. *Free will and luck*. New York: Oxford University Press.
63. Mele, Al. 2008. Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics* 12: 263–286.
64. Fischer, John Martin. 1994. *The metaphysics of free will: An essay on control*. Oxford: Blackwell.
65. Dennett, Daniel C. 2012. Review of *Against moral responsibility* by Bruce Waller and Dennett's rejoinder to Clark. Naturalism.org: http://www.naturalism.org/resources/book-reviews/dennett-review-of-against-moral-responsibility
66. Clark, Tom. 2012. *Clark comments on Dennett's review of against moral responsibility*. Org: Naturalism http://www.naturalism.org/resources/book-reviews/exchange-on-wallers-against-moral-responsibility#toc-clark-comments-on-dennetts-review-of-against-moral-respo-UN0XiqrR.
67. Pereboom, D. 2013. Free will skepticism and criminal punishment. In *The future of punishment*, ed. Thomas Nadelhoffer, 49–78. New York: Oxford University Press.
68. Sen, Amartya. 1985. *Commodities and capabilities*. North-Holland.
69. Sen, Amartya. 1999. *Development as freedom*. New York: Oxford University Press.
70. Nussbaum, Martha. 2011. *Creating capabilities: The human development approach*. Harvard University Press.
71. Powers, Madison, and Ruth Faden. 2006. *Social justice: The moral foundations of public health and health policy*. New York: Oxford University Press.