

REJECTING RETRIBUTIVISM: FREE WILL, PUNISHMENT, AND CRIMINAL JUSTICE, A PRÉCIS

GREGG D. CARUSO

SUNY Corning and Macquarie University

Keywords: free will, moral responsibility, punishment, criminal justice, retributivism, quarantine, public health, free will scepticism, desert, capabilities, hard incompatibilism, luck, social determinants, health, criminal behaviour, social justice

Within criminal justice systems, 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.

The dual aims of *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (2021) are to argue against retributivism and to develop and defend a viable nonretributive alternative for addressing criminal behavior that is both ethically defensible and practically workable. In the first half of the book, I develop six distinct arguments 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 the claim, which retributivism relies upon, that offenders are *blameworthy*

Contact: Gregg D. Caruso <gcaruso@corning-cc.edu>

 <https://orcid.org/0000-0002-9981-4922>

and *deserving* of punishment. I also consider a number of alternatives to retributivism, including consequentialist deterrence theories, educational theories, and communicative theories, and argue that they have ethical problems of their own. In the second half of the book, I then develop and defend my novel nonretributive approach, which I call the *public health-quarantine model*. The model draws on the public health framework and prioritizes prevention and social justice. I argue 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 nonretributive alternatives.

Along the way, I also explore the relationship between free will and criminal law, identify and document the social determinants of criminal behavior and argue that they are analogous to the social determinants of health, offer a number of specific policy proposals and prescriptions for implementing a public health approach to crime prevention, and defend a *capabilities approach* to social justice (see, e.g., Nussbaum 2011; Sen 1985, 1999), arguing that it can serve as the moral foundation of my public health framework and is consistent with my *free-will skepticism*, which maintains that who we are and what we do are ultimately the result of factors beyond our control (whether those be determinism, indeterminism, or luck), and because of this we are never morally responsible in the *basic-desert* sense.¹

My first argument against retributivism, which I call the *Skeptical Argument*, maintains that free-will skepticism undermines the retributivist notion that wrongdoers *deserve* to be punished in the backward-looking sense required. It argues that free-will skepticism is the only reasonable position to adopt when it comes to the problem of free will. Since retributive punishment requires the kind of free will associated with basic-desert moral responsibility in order to be justified, free-will skepticism implies that retributive punishment lacks justification.

1. Derk Pereboom defines *basic desert* moral responsibility as follows:

‘For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary. The desert at issue here is basic in the sense that the agent would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations’.

(Pereboom 2014: 2)

Understood this way, free will is a kind of power or ability an agent must possess in order to justify certain kinds of desert-based judgments, attitudes, or treatments—such as resentment, indignation, moral anger, and retributive punishment—in response to decisions or actions that the agent performed or failed to perform. These reactions would be justified on purely backward-looking grounds—that is what makes them basic—and would not appeal to consequentialist or forward-looking considerations, such as future protection, future reconciliation, or future moral formation (see Pereboom 2001, 2014; Levy 2011; Caruso 2021; and Caruso and Morris 2017).

Hence, insofar as we demand *justified* legal punishment practices, I maintain that we should reject retributivism in light of the philosophical arguments against free will and basic-desert moral responsibility. In support of free-will skepticism, I offer two distinct sets of arguments: one leading to a form of *hard incompatibilism* and the other *hard luck*. The first features distinct arguments that target the three leading rival views about free will—event-causal libertarianism, agent-causal libertarianism, and compatibilism—and then claims the skeptical position is the only defensible position that remains standing.² It maintains that free will is incompatible with *both* determinism and indeterminism (see, e.g., Pereboom 2001, 2014; Caruso 2012, 2021). The second maintains that regardless of the causal structure of the universe, we lack free will and basic-desert moral responsibility because free will is incompatible with the pervasiveness of *luck* (see Levy 2011; Caruso 2021).

My second argument against retributivism maintains that *even if* one is not convinced by the arguments against free will and basic-desert moral responsibility, it remains unclear whether retributive punishment is justified. This is because the burden of proof lies on those who want to inflict intentional 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. The problem, I maintain, is that all extant accounts of basic-desert moral responsibility fail to satisfy the high burden of proof required. I call this second argument the *Epistemic Argument* and it runs as follows:³

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 retributivist justification for legal punishment assumes that agents are morally responsible in the basic-desert sense and hence justly deserve

2. *Determinism* is the thesis that facts about the remote past in conjunction with the laws of nature entail that there is only one unique future. *Libertarian* theories of free will reject the thesis of determinism and defend an indeterminist conception of free will in order to save what they maintain are necessary conditions for free will: the *ability to do otherwise* in exactly the same set of conditions and/or the idea that we remain, in some important sense, the *ultimate source/originator* of action. *Compatibilism*, on the other hand, defends a conception of free will that aims to reconcile free will with causal determinism. It maintains that what's required for free will is not the falsity of determinism but that actions are voluntary, free from constraint and compulsion, and caused in the appropriate way.

3. See also Pereboom (2001), Vilhauer (2009, 2012), and Jeppson (2021) who offer similar arguments.

- to suffer for the wrongs they have done in a backward-looking, nonconsequentialist sense.
3. The justification for the claim that agents are morally responsible in the basic-desert sense provided by both libertarian and compatibilist accounts faces powerful and unresolved objections and as a result fall far short of the high epistemic bar needed to justify such harms.
 4. 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.

The four remaining arguments I develop against retributivism 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 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.

After a transitional chapter that examines consequentialist deterrence theories, educational theories, and communicative theories and argues that they have

ethical problems of their own that are difficult to overcome, the remainder of the book is focused on developing and defending my novel nonretributive and non-punitive alternative for addressing criminal behavior: the public health-quarantine model. The core idea of the model 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. Yet the model does 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. In fact, the model is completely nonpunitive and requires special attention to the well-being and dignity of criminals that would change much of current policy. Perhaps most importantly, the model also develops a public health approach that prioritizes prevention and social justice and aims at identifying and taking action on the social determinants of health and criminal behavior.

Public Health-Quarantine Model

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 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 their 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 them as to quarantine the nonresponsible 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, Pereboom and I resist this strategy (Pereboom and Caruso 2018; Caruso 2021). Instead, we maintain that incapacitation of the seriously dangerous is justified on the grounds 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 and 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, the model requires that we adopt the *principle of least infringement*, which holds that the least restrictive measures should be taken to protect public health and safety. 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.

Second, the quarantine model places several constraints on the treatment of criminals. 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). We do not, for instance, quarantine people for the common cold even though it has the potential to cause you some harm. Rather, we restrict the use of quarantine to a narrowly prescribed set of cases. Analogously, on this model, the use of incapacitation should be limited to only those cases where offenders are a serious threat to public safety and no less restrictive measures were available. 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 (156). Rehabilitation and reintegration would therefore replace punishment as the focus of the criminal justice system. Lastly, if a criminal cannot be rehabilitated and our safety requires their indefinite confinement, this account provides no justification for making their life more miserable than would be required to guard against the danger they pose (156).

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. Most importantly, it situates the quarantine analogy within the broader justificatory framework of *public health ethics*. 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, exposure to abuse and violence, poor environmental health, and addiction.

Since the *social determinants of health* and the *social determinants of criminal behavior* are broadly similar, or so I argue, the best way to protect public health and safety is to adopt a public health approach for identifying and taking action on these shared social determinants. Such an approach requires investigating how social inequities and systemic injustices affect health outcomes and criminal behavior, how poverty affects health and incarceration rates, how offenders often have preexisting medical conditions that include 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 social determinants of health if we want to improve health outcomes, it is equally important to identify and address the social determinants of criminal behavior.

Furthermore, the public health framework sees *social justice* as a foundational cornerstone to public health and safety (Powers and Faden 2006). 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.

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 2021).⁴ 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’ (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)

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

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.

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.

References

- 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. 2021. *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (New York: Cambridge University Press)
- Caruso, Gregg D. 2016. 'Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model', *Southwest Philosophical Review*, 32: 25–48
- Caruso, Gregg D., and Stephen G. Morris. 2017. 'Compatibilism and Retributive Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance', *Erkenntnis*, 82: 837–55
- Jeppson, Sofia. 2021. 'Retributivism, Justification and Credence: The epistemic Argument Revisited', *Neuroethics*, 14: 177–90
- Kelly, Erin. 2018. *The Limits of Blame* (Cambridge, MA: Harvard University Press)
- Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility* (New York: Oxford University Press)
- Nussbaum, Martha. 2011. *Creating Capabilities: The Human Development Approach* (Cambridge, MA: Belknap 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*, ed. by Gregg D. Caruso and Owen Flanagan (New York: Oxford University Press), pp. 193–222
- Powers, Madison, and Ruth Faden. 2006. *Social Justice: The Moral Foundations of Public Health and Health Policy* (New York: Oxford University Press)
- Sen, Amartya. 1985. *Commodities and Capabilities* (New York: Oxford University Press)
- Sen, Amartya. 1999. *Development as Freedom* (New York: Oxford University Press)
- Tabandeh, Armin, Paolo Gardoni, and Colleen Murphy. 2017. 'A Reliability-Based Capabilities Approach', *Risk Analysis*, 38: 410–24
- Vilhauer, Benjamin. 2009. 'Free Will and Reasonable Doubt', *American Philosophical Quarterly*, 46: 131–40
- Vilhauer, Benjamin. 2012. 'Taking Free Will Skepticism Seriously', *Philosophical Quarterly*, 62: 833–52