

Précis of *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*

Gregg D Caruso

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

The dual aims of *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*¹ are to argue against retributivism and to develop and defend a viable non-retributive alternative for addressing criminal behaviour 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 it. 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 non-retributive 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 behaviour, one that is superior to both retributivism and other leading non-retributive alternatives.

Along the way, I also explore the relationship between free will and criminal law; identify and document the social determinants of criminal behaviour 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, arguing that it can serve as the moral foundation of my public health framework as well as being consistent with my free will skepticism – which maintains that who we are and what we do is 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.²

1. GD Caruso, *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice* (Cambridge University Press 2021).

2. 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' (D Pereboom, *Free Will, Agency, and Meaning in Life* (Oxford University Press 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 D Pereboom, *Living Without Free Will*

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. Hence, in so far 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 – event-causal libertarianism, agent-causal libertarianism and compatibilism – and then claims the skeptical position is the only defensible position that remains standing. It maintains that free will is incompatible with both determinism and indeterminism. 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.³

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: (a) 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 behaviour is seriously wrong; (b) the retributivist 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. The problem, however, is that (c) 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 falls far short of the high epistemic bar needed to justify such harms; hence (d) retributive legal punishment is unjustified and the harms it causes are *prima facie* seriously wrong.

The four remaining arguments I develop against retributivism are independent of worries over free will and basic desert moral responsibility. They include the

(Cambridge University Press 2001); Pereboom, *Free Will, Agency, and Meaning in Life*; N Levy, *Hard Luck: How Luck Undermines Free Will and Moral Responsibility* (Oxford University Press 2011); Caruso, *Rejecting Retributivism*; and GD Caruso and S Morris, 'Compatibilism and Retributive Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance' (2017) 82 *Erkenntnis* 837–55).

3. See Caruso, *Rejecting Retributivism*, ch 2, for my arguments. See also Pereboom, *Living Without Free Will*; Pereboom, *Free Will, Agency, and Meaning in Life*; and Levy, *Hard Luck*.

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.⁴ 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 non-retributive and non-punitive alternative for addressing criminal behaviour: the public health-quarantine model. The core idea of the model is that the right to harm in self-defence and defence 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 non-punitive 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 behaviour.

1 PUBLIC HEALTH-QUARANTINE MODEL

The public health-quarantine model begins with Derk Pereboom's famous quarantine analogy, which draws on a comparison between treatment of dangerous criminals and treatment of carriers of dangerous diseases.⁵ In its simplest form, it can be stated as

4. See E Kelly, *The Limits of Blame: Rethinking Punishment and Responsibility* (Harvard University Press 2018).

5. See Pereboom, *Living Without Free Will*; Pereboom, *Free Will, Agency, and Meaning in Life*; GD Caruso, 'Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model' (2016) 32 *Southwest Philosophical Review* 25–48; D Pereboom and GD Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life', in

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.

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.⁶ Instead, we maintain that incapacitation of the seriously dangerous is justified on the ground of the right to harm in self-defence and defence 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.⁷ 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 are 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.⁸ Rehabilitation and reintegration would therefore replace punishment as

O Flanagan and GD Caruso (eds), *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience* (Oxford University Press 2018); GD Caruso and D Pereboom, 'A Non-Punitive Alternative to Punishment', in F Focquaert, B Waller and E Shaw (eds), *Routledge Handbook of the Philosophy and Science of Punishment* (Routledge 2020).

6. See Pereboom and Caruso, 'Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in life'.

7. Pereboom, *Free Will, Agency, and Meaning in Life*, 156.

8. *ibid.*

the focus of the criminal justice system. Lastly, if a criminal cannot be rehabilitated 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.⁹

In addition to these restrictions on harsh and unnecessary treatment, the model also advocates for a broader approach to criminal behaviour 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 behaviour 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 behaviour 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 behaviour, how poverty affects health and incarceration rates, how offenders often have pre-existing medical conditions including 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 behaviour.

Furthermore, the public health framework sees social justice as a foundational cornerstone to public health and safety.¹⁰ 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 behaviour 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

9. *ibid.*

10. See M Powers and R Faden, *Social Justice: The Moral Foundations of Public Health and Health Policy* (Oxford University Press 2006).

racism, sexism, poverty and systemic disadvantage as serious threats to public safety and it prioritizes the reduction of such inequalities.

The core of the public health-quarantine model, then, is that the right to harm in self-defence and defence 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 behaviour. This combined approach to dealing with criminal behaviour, I argue, 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.