The Quarantine Model and its Limits

Andrea Lavazza[1,2] · Sergei Levin[3] · Mirko Farina[4,5,6]

Abstract

There are several well-established theories of criminal punishment and of its justi-
fication. The quarantine model (advocated by Pereboom and Caruso) has recently
emerged as one of the most prominent theories in the field, by denying the very idea
of criminal justice. This theory claims that no one ought to be criminally punished
because fundamentally people do not deserve any kind of punishment. On these
grounds, the quarantine model proposes forms of incapacitation based on public
safety considerations. In this article, we briefly review a series of objections raised
against the quarantine model and propose some new or revised arguments, which are
aimed at showing its inconsistencies and weaknesses. These arguments are related to
(a) the lack of a reliable way of determining who is dangerous and the consequent
need to make judgments about confinement based on probabilities, and (b) the pros-
pect that the quarantine model may encourage certain crimes. Given the arguments
we present in this paper, the quarantine model proves to be less solid, humane, and
desirable than its proponents claim.

✉ Andrea Lavazza
    lavazza67@gmail.com

    Sergei Levin
    levins@tcd.ie

    Mirko Farina
    m.farina@xmu.edu.cn

1   University of Pavia, Department of Behavioral Sciences, Piazza Botta 1 Pavia, Italy

2   Centro Universitario Internazionale, Arezzo, Italy

3   Trinity College Dublin, Trinity Business School, Dublin, Ireland

4   Institute for Digital Economy & Artificial Systems [IDEAS], Xiamen University, 15/F,
     16/F, Building B09, Golden Brick Future Innovation Park, Jimei District, Xiamen,
     Fujian Province 361021, People's Republic of China

5   Institute for Digital Economy & Artificial Systems [IDEAS], Lomonosov Moscow State
     University [MSU], Lomonosov Building, Lomonosovskiy Prospekt 27, Ramenki District,
     Moscow 119192, Russian Federation

6   University of Religions and Denominations, Qom, Iran

1 Introduction

In legal practice and among laypeople the concepts of retribution and incarceration of wrongdoers are taken as the cornerstones of modern theories of punishment, which is often seen as an inherently retributive practice (Bedau & Kelly, 2019).

Retributivism (Husak, 2013; Kershnar, 2000; Moore, 1997; Murphy, 2007; Von Hirsch, 1992), is a theory of justification of punishment, according to which "a person who unjustifiably and inexcusably causes or risks harm to others or to significant social interests deserves to suffer for that choice" (Berman, 2008, p.269) and the suffering deserved is proportionate to the violation the offender has committed.

The three main principles underlying retributivist theories of justice are: (a) those who commit acts deemed wrongful morally deserve to suffer a proportionate punishment; (b) it is morally good that the offender is punished as he deserves by a legitimate punisher; (c) it is not morally permissible to intentionally punish an innocent person or excessively punish an offender (Walen, 2020). The law provides the framework for the justification of a single act of punishment, while the justification of the practice itself may be found in a consistent theory, a set of shared values, or in some goals that society, as a whole, aims to pursue.

Retributivism is based on the concepts of positive and negative desert (Tadros, 2017). The positive desert claim holds that offenders morally deserve punishment for their wrongful acts (and, in a weaker version, that there is some intrinsic positive value in punishing an offender). The negative desert claim maintains that it is not morally acceptable to punish those who have not committed any wrongful act. Overall, it can be said that retributivism is a deontological theory that is essentially backward-looking.

Retributivism is often contrasted, sometimes with excessive simplification, with another theory of punishment, namely consequentialism, which is forward-looking. In general terms, consequentialism maintains that the right action depends on the consequences of that action, and, concerning wrongful acts, seeks "to maximize good consequences of punishment such as crime deterrence, incapacitation of dangerous people, and rehabilitation of offenders" (Kolber, 2018, p.517).

Nowadays, many judicial systems are mixed; that is, they rely on both backward-looking retributive justification and forward-looking instrumental justifications such as public safety and social order. From a theoretical perspective, retributivism has been criticized for being excessively harsh or for imposing unnecessary suffering upon people for deeds they perhaps committed but might not be responsible for (e.g., Shafer-Landau, 1996). Consequently, the main assumptions of retributivism have come under scrutiny, especially the idea that the offender always deserves to be punished and that, from the epistemological point of view, we know with reasonable accuracy which punishment the offender

deserves (Caruso, 2020). Central to this debate is the concept of *basic desert*, a notion suggesting that agents inherently deserve rewards or punishments based solely on their actions or character, irrespective of consequentialist considerations. This concept has been particularly criticized for its primary assumption that agents possess free will. Such possession is foundational for many retributive theories, as it posits an agent as ultimately responsible for their actions (Caruso, 2021; Caruso & Dennett, 2021; Levy, 2011a; Pereboom, 2001, 2014).

In this debate, a plea for a more compassionate and effective alternative to the punitive approach has emerged. Karl Menninger (1968) advocated for prioritizing understanding and rehabilitation over retribution, emphasizing that most criminals have treatable mental disorders. He was criticized by Jeffrie Murphy in the 1970s (1979). However, in recent years, the idea of perceiving offenders as part of the crime problem, rather than its source, has regained attention. This quarantine model was endorsed by Pereboom (2001, 2014) and later defended by Caruso (2016, 2021). The model aims to prioritize social justice and mitigate unnecessary suffering. It purports to provide a legitimate framework for justifying quarantine (and very light criminal sanctions) that is purportedly more humane than approaches informed by either retributivism or consequentialism. The fundamental philosophical underpinning of this model is free will skepticism, that is, the belief 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" (Pereboom & Caruso, 2018, 195). However, this does not negate other forms of responsibility. Thus, the foundational notion of the quarantine model is that no justification for punishment exists, and justifying quarantine does not equate to endorsing an alternative form of punishment.

The quarantine model is not immune from criticism and potential concerns. In this article, we will not try to refute the arguments that support free will skepticism, whether conceptual or empirical. Our aim, here, is more practical: we want to show that the quarantine model is not the most consistent, humane, and effective way of dealing with offenders and wrongdoers, despite its laudable objective of sparing people unjustified harm.

After reviewing some objections raised against the quarantine model, we propose (building and expanding on previously published work: Reference Suppressed) some new or revised arguments, which are aimed at showing the inconsistencies and weaknesses of the quarantine model without calling into question the premise of free will illusion (on which we remain agnostic here). These arguments are related to (a). the lack of a reliable way of determining who is dangerous and who is not, and (b). the possibility that the quarantine model may encourage certain crimes. Given these arguments, the quarantine model proves to be less solid, humane, and desirable than its proponents claim.

2 The Quarantine Model

There are, broadly speaking, two types of arguments developed by researchers to undermine the existence of free will. These are: (i) scientific arguments against free will based on Libet-like neuroscientific experiments (Libet et al., 1983; Harris,

2012; Cashmore, 2010) and empirical psychology experiments (Wegner, 2002), that purport to show that what we experience as our conscious will is not the real cause of our actions, since the latter starts before the former appears; and (ii) philosophical arguments for free will skepticism that have less to do with conscious control and more to do with determinism/indeterminism (Waller, 2011, 2015), *causa sui* (Strawson, 1994), manipulation (Pereboom, 2001, 2014), and luck (Levy, 2011b). Crucially, these types of arguments tend to assume that the agent is not the appropriate source of their actions, given the way the world is and the beings on it are.

Among these latter philosophical arguments, there are those specifically developed by Pereboom (2001, 2014) and Caruso (2016), who questioned whether it is ever rationally justified to hold someone morally responsible in the basic desert sense. Such philosophers have called for a significant reform of the penal and judicial system and explicitly endorsed a model of incapacitation for dealing with dangerous criminals that challenges retributivism, treats offenders as not responsible for their crimes, and envisages that they should be detained in humane conditions, while being offered the opportunity for rehabilitation. In this section, we summarize the crucial tenets underlying this model and look at some of the arguments proposed to support it.

The main tenet underlying the quarantine model is that the confinement of those who have proven to be dangerous for society is more humane, hence preferable, than alternatives currently used for dealing with criminal behavior. The model includes also preventive detention in the form of non-punitive quarantine with the goal of preventing future crimes. The detention should be implemented only after we have strong evidence of harmful tendencies. Typically, the triggering event for isolation should be that an individual has actually been convicted of having harmed another. Pereboom's theory of crime prevention is meant to create a social system capable of minimizing criminal activity without the need of imposing harsh punishments. Instead of focusing on retribution, Pereboom and Caruso rather look at measures to prevent crimes.

To justify the detention (both after crime commission and preventive isolation) of dangerous subjects, Pereboom draws an analogy with the process of quarantining infectious people. Quarantine is a process whereby a restraint is placed upon people for a certain time, in order to prevent the spread of contagious illnesses, diseases, or pests. Its aim is to prevent the spread of an infection, from the quarantined individual to other members of society. This separation can be voluntarily applied or forcefully imposed (by an external actor—the state, for example) on people who do not wish to be quarantined.

Pereboom identifies—at least—three major similarities between the process of quarantining people, described above, and its idea of preventive detention. The existence of these relevant similarities allows Pereboom to make a strong analogy between these two practices.

The first similarity is that both subjects (the contagious patient and the dangerous criminal) pose a serious threat to those around them and to society in general. As a carrier of the Ebola virus constantly threatens people while they wander around (potentially causing—by transmitting the infection- excessive pain, long-term suffering, and even death to people it encounters), so, does a potentially dangerous

criminal (think of a psychopath, for instance), who is continuously on the verge of committing a horrendous crime while they are free. Like carriers of diseases, violent offenders thus risk causing severe injuries, and even death, to those surrounding them. Thus, in Pereboom's view, both violent criminals and disease carriers threaten public safety.

The second similarity is that, in each case, both the contagious patient and the potential criminal cannot avoid harming others without the adoption of some restrictive measures. The measures should be applied to dangerous individuals in accordance with the principle of the least infringement (more on which below), which specifies that the least restrictive measures should be taken to protect public health and safety (Caruso, 2021, p. 292).

Lastly, and related to the second similarity discussed above, a detainee cannot be held responsible in a basic desert sense for the harm they may cause to others (Pereboom, 2014). This similarity can be substantiated only if we operate under the premise (as Pereboom does) that human agents do not possess free will and, therefore, are not free agents. In Pereboom's account, the offender is not the appropriate source of their actions in the sense needed for basic desert moral responsibility, as the disease carrier cannot avoid infecting others with their illness. As individuals cannot choose to become contagious carriers, so subjects are driven towards criminal behaviors, not by their own free will; rather by genetic dispositions, specific environments, and similar factors, which are all outside of their control. In this vein, the concept of luck and its pervasiveness is of primary relevance in undermining free will and basic desert (Caruso & Dennett, 2021; Levy, 2011a).

Pereboom thinks that, to protect society from the harms of crime, we should not punish offenders; rather, since they don't deserve to suffer for the harms they have caused, we should place criminal offenders in preventive detention as we would do with those quarantined with dangerous contagious diseases. However, how can an advocate of the quarantine model justify detention, separation, or isolation of non-free (therefore non-morally responsible in the basic desert sense) agents? Pereboom justifies detaining individuals on the grounds that quarantining them prevents society from injury, and, unlike retribution, does not depend upon blameworthiness, or free choice in general.[1]

In addition, the quarantine model envisages different levels of isolation, according to Pereboom. 'As less dangerous diseases justify only preventive measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate (e.g. just monitoring) restraints' (Pereboom, 2014, 156). Furthermore, incapacitation revolves around the idea of well-being and rehabilitation. 'Just as fairness recommends that we seek to cure the diseased we quarantine, so fairness would counsel that we attempt to rehabilitate the criminals we detain' (Pereboom, 2014, 156).

More recently, Pereboom's model has been further developed by Caruso, who has framed the quarantine model within a broader justificatory framework drawn from public health ethics. This new model is called the public health-quarantine model

---

[1] To the best of our knowledge, the first researcher who clearly drew this analogy was Schoeman (1979).

(Caruso, 2016, 2017, 2021). According to this author, it provides a framework for justifying quarantine and criminal sanctions that is more humane than retributivism, and preferable to other non-retributive alternatives in that it prioritizes prevention and social justice (Caruso, 2016, p. 4).

At the core of the public health-quarantine model, much the same as the quarantine model previously discussed, there is the idea that retributive punishment cannot be justified because offenders are not morally responsible for their own actions. The principle on which this model is based—the principle of least infringement—holds that the least restrictive measures ought to be taken to protect private (self-defense) and public health and safety (defense of others) (Pereboom & Caruso, 2018). Adopting this principle is meant to ensure that sanctions will always be proportionate to the danger posed by a given individual and will never be unjustified.

The public health ethics model, however, not only justifies quarantining to protect public health, but also actively demands that we take the necessary steps to prevent quarantining from occurring in the first place. This means that quarantine is the last possible measure available to us, when the public health system fails in its primary function (that is, to prevent harm). So, the public health-quarantine model shifts the focus of our attention from the penal/judicial system (the dimension of punishment and retribution) to the social one (the dimension of prevention), forcing us to identify and possibly tackle the systemic causes of crime.

More specifically, Caruso (2017), argues that the social determinants of health (SDH) and the social determinants of criminal behavior (SDCB) are quite similar. He therefore believes that we should adopt a broader public health approach for identifying and taking actions on these determinants. Related to this point are issues of social justice, which are seen as a cornerstone to public health and safety. Caruso's model tries to identify and remedy social and economic inequalities (such as racism, sexism, poverty, and other systemic disadvantages), which are considered to be triggering factors for crimes.

Pereboom embraced much of Caruso's emphasis on the need to consider both SDH and SDCB, essentially welcoming a partial variation in scope and focus of his original model. Yet, Pereboom and Caruso recently slightly disagreed on the type of responses given to critics (general deterrence theorists) of their view (see especially Pereboom, 2020). It is therefore worth briefly mentioning this difference as a point of divergence between Pereboom and Caruso. We note that this issue is discussed at length in Caruso's most recent book (2021: sec 9.2).

In response to considerations raised by Victor Tadros (especially 2017; see also Levin et al., 2021), Pereboom (2020) has conceded that perhaps a greater level of general deterrence might be desired than what is permitted by the right of self-defense. In particular, he has argued that 'with regard to monetary fines and short prison terms it might sometimes be justified to use unfree wrongdoers in ways that involve such penalties to subserve general deterrence' (quoted in Caruso, 2021: 310).

Caruso (2021: 310–315) agrees with Pereboom that 'incapacitation justified on the right of self-defense will naturally produce free general deterrence' … but 'contends that this is all we should seek to justify'. In other words, while Pereboom is open to ad hoc revisions of the quarantine model (e.g., adding punitive measures such as fines

and short-term sentences) to deter specific crimes, Caruso is not ready to make such a concession. As such his version of the public health–quarantine model remains pure and free of any punitive components and -Caruso claims (2021: p. 313)- is therefore 'compatible with the complete rejection of all justifications of punishment'.

Thus, summarizing these views, Pereboom and Caruso both believe 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 of a society and its members. Their models also focus on the well-being of criminals, on prevention and social justice, demanding us to take actions against the social determinants of health and criminal behavior. Pereboom and Caruso, however, disagree on some general issues related to deterrence.

In the next section, we will review some criticisms that have been raised against the quarantine model (and its more recent variant: the public health quarantine model) and look at how proponents of the quarantine model have defended themselves against such criticism.

3 Criticisms of the Quarantine Model (and Responses by its Proponents)

Michael Corrado, who denies basic desert moral responsibility but endorses the usual hard treatment of reasons-responsive criminals, has proposed two main objections against the account defended by Pereboom and Caruso. The first objection raised by Corrado concerns the alleged lack of distinction between different types of offenders. According to Corrado, Pereboom and Caruso treat all criminals through the illness model, failing to consider that there are criminals capable of responding to reason and criminals who are incapable of doing so. The second objection is that too many people would be drawn into the criminal justice system by the quarantine model. This is because anyone who presents even a potential danger to society becomes a candidate for incapacitation (Corrado, 2016, 2019).

Pereboom and Caruso have responded to Corrado's criticism. With respect to the first objection, involving lack of distinction between different types of offenders, they note that Pereboom (2014) already rejected the idea that all criminal tendencies are exclusively forms of psychological illness, modeled on physical illness. Pereboom and Caruso's view does consider the conditions that can lead an offender to criminal behavior; such conditions are not restricted to psychological illness, but also include 'insufficient sympathy for others or a strong tendency to assign blame to others and not to oneself when something goes wrong' (2018, p. 207–208). Pereboom and Caruso in general 'aim to bring about moral change in an offender by non-punitively addressing conditions that underlie criminal behavior' (2018, p. 208). Finally, Pereboom and Caruso themselves acknowledge potential ethical concerns with respect to treatments -like deep brain stimulations- that bypass an individual's rational capacity. They stress that 'rehabilitation methods that directly appeal to a criminal's rational capacities should always be preferred and attempted first' (2018, p. 213).

With respect to the second objection raised by Corrado—the idea that an excessive number of people would be drawn into the criminal justice system if we adopted the quarantine model—Pereboom and Caruso respond as follows. They claim that their view doesn't really prescribe that all dangerous people be detained until they are no longer dangerous. They instead affirm that some types of persistent threats can be addressed by monitoring, as opposed to detention. Also, they note that certain behaviors, despite being considered criminal, may not result in incapacitation. These are behaviors that are already being decriminalized today, such as those related to the recreational use of drugs.

Concerning the incapacitation of dangerous people who have not yet committed a crime, Pereboom and Caruso's response is twofold. On the one hand, they point out that there are moral reasons that suggest using a lot of caution in this respect (Pereboom & Caruso, 2018). One must consider the right to liberty, the imperative not to use people as means, and the risk of abuse posed by the state's ability to make preventive arrests. Moreover, procedures to ascertain the potential danger posed by an individual are invasive and often unreliable (we elaborate this point in Sect. 4.1 below).

On the other hand, however, Pereboom and Caruso point out that tests may soon be available to detect neural factors that predict violent and dangerous behaviors (Pereboom & Caruso, 2018, 215). In those cases, their model could endorse preventive detention. However, they argue, this should not be seen as a strong objection to their theory because, in their opinion, everyone would agree to arrest a person who will commit a serious crime within a few days, if there were no alternatives to stop them.

Another important line of criticism was recently voiced by Smilansky (2019).[2] He contends that deontological constraints, which Pereboom claims to endorse, in order to differentiate his view from utilitarianism (that most critics claim that innocent people could be punished, if it turns out to be helpful), are very weak if one detaches them from their basis. Their basis, however, seems to be precisely the belief in free will, responsibility, and autonomy of the agent; something Pereboom straightforwardly denies. Following Smilansky, we believe that the deontological constraints to spare innocent people from unjust punishment are built upon concepts that the advocates of the quarantine model are not prepared to endorse.[3] One way for the free will

---

[2] Lemos (2016, 2018 ch. 6, 2019) claims that the quarantine model could justify the use of lower evidentiary standards and that it lacks theoretical grounds for the kind of dignity which can be considered fundamental to humans. Pereboom and Caruso replied that they justify incapacitation based on the right to self-defense and defense of others, and this right does not extend to those who are not posing a threat.

[3] Caruso (2021: 200–202) argues that certain deontological constraints are justified within his model and can be reconciled with free will skepticism. He maintains that it is possible to have a non-desertist theory of justice and provides as an example the theories of John Rawls and David Hume. However, Smilansky's challenge was not that only deserts theories of justice are possible; rather that without believing in free will one loses the Kantian motivation not to treat other people as a means for achieving better consequences. John Rawls himself suggested a hybrid theory of punishment that allows retributivist sentencing to achieve utilitarian goals (Rawls 1955) and one can raise the objection to his theory that it allows inflicting suffering on the criminals for the sake of protecting society. Hume's theory is also not a very good example to prove the point. One the one hand Hume was not a free will skeptic he was rather a compatibilist. One the other hand his forward-looking account of responsibility is open to the deontological 'use' objection (Russell 2021).

skeptic to address the objection raised by Smilansky was suggested by Benjamin Vilhauer (2013). Rather than merely rejecting action-based desert, they could advocate for personhood-based desert. Vilhauer contends that personhood-based desert, which maintains Kant's regard for persons, is distinct from action-based desert. He further posits that while skepticism about free will calls into question action-based desert, it does not negate personhood-based desert.[4]

4 New Objections to the Quarantine Model

As we have seen above, Pereboom and Caruso have answered most of the objections that have been raised against their model; however, one could plausibly think that the jury is still out on whether they have done so effectively. In what remains of this article, we introduce a series of new objections or reframed ones against the quarantine model that, taken individually or together, may point to further inconsistencies and weaknesses in Pereboom and Caruso's account.

4.1 Practical and Ethical Issues with Predicting Recidivism

In this section, we consider how the quarantine model addresses the issue of predicting recidivism. We argue that Pereboom and Caruso, at the time of writing, do not provide sufficient details to address this potential issue. First, we point out that the criminal justice system does not have the effective tools to predict recidivism with reasonable accuracy. Second, pace Pereboom and Caruso, we show that the probabilistic approach to recidivism they suggest embracing raises some important ethical concerns.

One of the fundamental tenets underlying the quarantine model is that it justifies incapacitating the criminally dangerous with the minimum harm required for adequate societal and individual protection (Pereboom & Caruso, 2018, 207). Pereboom and Caruso claim that offenders cannot be held as morally responsible, in the backward-looking sense; thus, they claim that the offender should be incapacitated only for the time required for the protection of others.

This approach supposedly minimizes potential future harm to both the possible victims and the offender. If an offender poses no risk, then there is no need to isolate them (in Sect. 4.2, we will consider some exceptions that Pereboom discusses). In other words, the quarantine model states that no one should be in prison if we do not reasonably expect that they may commit any new crimes. That would allow society to spare financial resources and it would be more humane for the potential offenders, as they would not suffer more than needed for the sake of public safety.

This feature of the quarantine model can be praised and may be considered to be a better alternative to retributivism, as it might be said to prevent unjust punishment and favor rehabilitation. In addition, the quarantine model can be said to provide a rationale for prolonging preventive detention to protect society against

---

offenders, who are likely to commit new crimes. Current justice systems, in some cases at least, release criminals who served their sentences when they were still dangerous to society. Detaining dangerous offenders for as long as needed, for rehabilitating them, could indeed prevent serious harm and societal suffering in the long-term.

So, the quarantine model promises two desirable results. First, it is poised to eliminate needless detention of offenders who do not pose a threat to society; second, it seems to be able to reduce serious societal harms by prolonging the quarantine of dangerous offenders. However, to achieve these results, one needs to be able to assess, with high accuracy, the level of future criminal readiness of possible offenders. This is crucial for the quarantine model: without a reliable system of assessment, criminals could spend too much time in prisons or too little. If the former is the case, then the quarantine model would end up being unnecessarily cruel; if the latter is the case, then the quarantine model would fail to be effective for self-defense against criminals.

The analogy with the quarantine process in public health is supposed to be unproblematic. We need to separate carriers of contagious diseases for the time they are dangerous or potentially dangerous to other individuals. Luckily, medical knowledge and diagnostics tools allow doctors to make well-informed and effective decisions in this respect. People who have been exposed to the virus need to be quarantined to see if they became sick (Gordis, 2014, 28). The duration of the quarantine is determined by the incubation period of the disease (for instance, less than 14 days in the case of COVID-2019) with some extra time usually allowed, just to be completely sure that the patient is virus-free. We know that people pose no threat after the end of the incubation period. The quarantine may end even faster if the tests show that the person is not infected.

True, in clinical decision-making doctors work with diagnostic probabilities (Gillies, 2018; Sox, 1986). However, the likelihood of someone being infected may be safely established for many diseases. Moreover, it is worth noting that with multiple testing available diagnoses tend to be very accurate. All in all, there are many reliable ways in which doctors could know, with sufficient accuracy, whether a given patient is virus-free; that is, whether she is no longer dangerous and no longer requires isolation. They may check the patient's clinical status, run special medical tests, or just wait for the incubation period to pass. So, for actual or suspected carriers of contagious diseases, there is a reliable mechanism in place for determining their time of safe release. This means that a medical decision to lift restrictions for specific individuals is based on reliable predictions (although medicine and biology are not exact sciences and there often are some exceptions to the rule).

The level of accuracy that accompanies a judicial decision to release a person from social quarantine is nowhere near the level of accuracy that can be achieved by medical professionals releasing someone from quarantine. As such, it can, at best, be based on averages. Of course, there would be some cases where a crime could be considered imminent, beyond any reasonable doubt. For instance, a terrorist who shows no remorse and openly proclaims that he is prepared to continue his fight if released, or an offender with a mental disorder, that is the cause of persistent violent behavior.

However, for many wrongdoers, their future dangerousness is not reliably predictable. If, for many offenders, there is no way to reliably assess whether the person could re-offend, then the quarantine model is not in a better position than retributivism with respect to determining the right time for detention. Pereboom and Caruso blame retributivists for not providing precise guidelines concerning the limitations of punishment for violent crimes. However, assessing the future dangerousness of an individual in practice is not less difficult than determining the length of detention based on the principle of proportionality on which retributivists rely.

Some convicted criminals say that they regret what they have done and promise they would never do it again. Later, some of them will re-offend and some of them won't. How could we reliably decide whether preventive detention could be justified in each case? Members of parole boards and judges struggle with this question all the time and their decisions are far from perfect (Henry, 2021; Laqueur & Copus, 2022). No test or textbook in criminology can tell us for how long a certain criminal would need to be incapacitated in order to be rehabilitated. Thus, unlike the cases discussed above, there are no reliable tests to use, to determine whether the offender would commit a new crime in the future. As we mentioned earlier on, Pereboom and Caruso may object that neurological tests for violent tendencies will be developed in the near future and that such tests may allow us to determine, with reasonable accuracy, whether someone is likely to commit violent crimes (Pereboom & Caruso, 2018, 215). However, we do not have such tests at the moment, and there is no guarantee that they will ever be developed.

Proponents of the quarantine model may concede that criminal predictions lack the level of accuracy achieved by medical predictions, with respect to infectious disease, but could reply that no other penal system is able to do that. However, this problem seems to mostly affect the quarantine model as, for example, retributivists are not necessarily committed to assessing the future dangerousness of offenders; for they are content to assess whether a wrongdoer deserves to be punished in proportion to the crime he committed.

As a solution to this problem, the advocates of the quarantine model could suggest introducing risk-assessment tools, to estimate the probability that a wrongdoer could re-offend. Risk-assessment methods are already widely used (Singh et al., 2014). Such methods employ computer algorithms, big data analysis, and advanced statistical calculations to assign a probability of re-offending to an individual (wrongdoer) based on the risk factors associated with their profile.

There are numerous risk assessment tools, for example, the Risk-Need-Responsivity (RNR) model, which is perhaps the most widely known model for the assessment and treatment of offenders (Ward et al., 2007). The RNR model can be taken to be based on the personality and cognitive social learning theory of criminal conduct. Other risk assessment tools are based on statistical approaches, which showed a greater accuracy than clinical predictions made by mental health practitioners (Ægisdóttir et al., 2006; Campbell et al., 2009). A significant blow to the reputation of risk assessment tools was given by a recent empirical study that found that big data and advanced machine learning, as well as widely used commercial risk assessment software based on statistical approaches (such as COMPAS), are no more accurate or fair than predictions made by people with little or no criminal justice

expertise (Dressel & Farid, 2018). The latest generation of these risk-assessment tools is more promising. They use "mobile, self-scoring, risk assessment software that relies on neurocognitive testing to predict re-offence. Results gave a recidivism prediction value of 0.70" (Haarsma et al., 2020). Yet, none of the current risk-assessment tools that have been developed has reached the threshold where the predictions are credible beyond a reasonable doubt, akin to the diagnostic accuracy we expect when releasing someone from medical quarantine.

However, the usage of predictive tools in the court is on the rise (Raaijmakers, 2019; Reiling, 2020; Xu, 2022). There is also a growing literature that points to automated systems as an antidote against the noise that often characterizes human decision making in such contexts (Kahneman et al., 2021). Yet, evidence also points out that the application of predictive tools in law enforcement risks introducing discriminatory measures such as racism, sexism, and other social biases (Angwin et al., 2022; Crawford, 2021).[5] There is therefore a significant margin of moral arbitrariness involved in such decisions as well as about the type of error one wants to minimize (freeing a dangerous criminal or keeping a non-dangerous citizen in quarantine).

In addition, there is an ongoing debate over the fairness, applicability and efficiency of risk assessment procedures in the criminal justice system (Hamilton, 2015; Harcourt, 2015; Skeem & Lowenkamp, 2016; Slobogin, 2021; Starr, 2014). For example, a recent study points out that when it comes to risk assessment "it is impossible to maximize accuracy and fairness at the same time and impossible simultaneously to satisfy all kinds of fairness" (Berk et al., 2018). Pereboom and Caruso should acknowledge the existence of this ongoing debate as well as its consequences, if the quarantine model is to resort to risk-assessment (and we see no way how it can avoid it).

Here, moreover, we just stress one ethical concern, that we believe is central to risk-based detention strategies. The probabilistic nature of risk-assessment would inevitably lead to the incarceration of some individuals who would not re-offend. Huemer (2010) and Aboodi et al. (2008) use the criminal justice system as an example of a system in which an accumulation of individually risky (but not too risky) actions may become statistically significant to indicate the risk of future wrongdoings. Denise Meyerson provides a more precise description of the issue:

> Suppose it is said that the likelihood of violent offenders re-offending is 0.75. This means that three out of four of the group of violent offenders will re-offend. Now let us suppose that 400 violent offenders who have served their sentences are about to be released from prison. If they are all preventively incarcerated on the basis that three in four of them will re-offend, 100 harmless individuals — false positives — will be imprisoned unnecessarily (Meyerson, 2009, 510).

---

[5] As it encourages the usage of these tools, this could well be another reason for worrying about the quarantine model.

Detaining people based only on probability means that we are prepared to limit the liberty of some non-dangerous people for the sake of public safety, which sounds very consequentialist.

To that objection, Pereboom and Caruso may reply that the detention of people, based on statistical generalization, is still justified for self-defense, as it brings overall better consequences than any alternative. To be more specific, they can suggest a proportionality in assessment. If the harm caused by the crimes, discounted by the probability, is greater than the harm caused by detention, then detain; otherwise release. Thus, such an approach seems to justify detention on consequentialist grounds, which Pereboom and Caruso however reject.[6] Even if they decide to adopt a consequentialist justification, the great variety in kinds of crimes makes it -we believe- difficult to follow such a path. For minor offenses, it may work, but it is hard -if not impossible- to compare non-financial harms for serious crimes (murder, rape, torture etc.) and harms caused by detention in modern civilized detention facilities. It seems like repeatedly adding a few extra years of detention is almost always a lesser evil than even a small chance for recidivism of the most heinous crimes.

To avoid the problem of indefinite detention, Caruso recommends "placing the burden of proof on the state to establish, at regular intervals, that the threat posed by an offender warrants continued incapacitation" (2021, p. 325) and for parole boards to consider expert advice on the matter. However, we already noted that parole boards lack tools for accurate criminal prediction and their decisions are often far from perfect. If Caruso is suggesting a mechanism that hasn't proven reliable, it would be fair to ask him for further explanation on how he thinks we may improve its functioning. Another possible reply by Pereboom and Caruso could be to allow releases earlier than absolute safeness is reached. This might be done to prioritize individual liberty. Pereboom and Caruso should explicitly acknowledge this trade-off if they agree to it. This would explain their stance on the matter while also highlighting the unavoidable imperfections of the quarantine model. Nevertheless, addressing risk assessment in their writings remains essential, and we hope that our paper will push them to look more seriously into this matter.

The concerns that we raised in this section may be partially motivated by the lack of details provided by Pereboom and Caruso, regarding the way preventive detention should be implemented. The authors could claim that our concerns about the prediction of reoffending are too technical, and it is a criminologists' job to address them. However, preventive detention is justified, and the quarantine model becomes more humane than alternative criminal models only if it is possible to detain the individual for the right amount of time.

At the time of writing, there is no effective mechanism in place to reliably determine how much time is fair and sufficient to prevent reoffending in each case, as shown by the fact that scholars have been extensively discussing the ethical issues related to risk assessment and by the constant struggles that judges and members of parole boards have to face in daily cases.

---

[6] Although Pereboom agrees to take into account the consequences of moral and legal decisions (2018).

In the next section, we will present another criticism that stems from the problem of risk-assessment, but it is specific to the quarantine model since it does not include deterrence in the classic sense.

4.2 The Quarantine Model Could Encourage Specific Crimes

Echoing Smilansky (2017) but extending his criticism to focus on one-time offenders, we argue that the quarantine model could facilitate non-recursive crimes. Our criticism is that those who committed a crime and pose no further threat, if we were to adopt that model, could suffer very minor consequences—even for a very serious single offense. That may motivate people to commit crimes which, under standard retributivist or consequentialist conditions, would instead be at least deterred from being committed.

There are four possible combinations of innocence and dangerousness that the quarantine model must address.

(1) Innocent person who does not pose a threat to society
(2) Innocent person who poses a threat to society
(3) Offender who poses a threat to society
(4) Offender who does not pose a threat to society

By 'innocent', we meant an individual that did not commit any criminal act (leaving aside questions of moral responsibility). (1) This category is easily dealt with by the quarantine model: we must never incapacitate such people. (2) Pereboom and Caruso provide consistent reasons for incapacitating such people if the threat they pose is evident beyond a reasonable doubt. (3) The quarantine model explains that we should incapacitate dangerous offenders based on our right to self-defence. (4) This is where Pereboom and Caruso do not explore the full range of possibilities. Pereboom and Caruso discuss offenders who committed petty crimes and suggest society would be better off without putting them in prison. That seems a reasonable thing. However, there is a logical possibility that there could be offenders who committed a serious crime yet pose no future threat to society. We argue that the quarantine model has no moral grounds for their incapacitation. As a consequence, the quarantine model lacks adequate deterrence for (4).

In the previous section, we established that the risk-assessment procedures must be at the disposal of police and those who will be in charge of dealing with potential criminals for a quarantine model to work. If a quarantine model were to be implemented in the future, the general public would know about these assessment-procedures. The criteria used could not be kept secret, because, in a democracy, the reasons or grounds for which a person is confined must be public and openly debated. Pereboom writes that people should know what the state does to its members, and why, when someone is dangerous to others. Moreover, the knowledge about the implemented system of quarantine may serve as a general deterrence (Pereboom, 2019, 103). Thus, people would know that the criminal justice system would focus on prophylactic measures and if those measures failed, offenders would be subjected

to rehabilitation and incapacitation. The intensity of rehabilitation and the period of isolation would depend on the risk of future harm posed by a potential one-time offender.

Let's imagine one country implemented the quarantine model. And suppose that someone named Lora lives there. Lora desires to murder her mother, Klara out of hatred. Klara raised Lora alone. She was a controlling mother and had lots of highly irritating habits (we leave the details to the reader's imagination). However, Klara never did anything that falls under the definition of child abuse or domestic violence. The relationships between Lora and Klara were increasingly difficult and over the years Lora acquired a strong desire to kill her mother.

Lora is neither mentally nor physically ill, even though she suffered adverse experiences in her childhood. The desire to kill her mother is not overwhelming; it is just very strong. Lora doesn't want to do it at any cost. She is not an intrinsically bad person, insofar as she just hates her mother. Klara is the only person whom Lora hates. Education, employment, and family support are considered protective factors that inhibit the likelihood of reoffending. Lora has a very low-risk score. She is over 45, does not have a criminal history, received a good education, has a stable full-time job, and has a functional family of her own. Nevertheless, the memory of her hardships and every interaction with her mother stirs up hatred with renewed vigor.

Lora knows that she lives in a society that endorses the quarantine model. One day, Lora decides to seize the opportunity to take revenge on her mother. To her delight, Lora understands that if she murders her mother, her individual probability of reoffending will turn out to be negligible.[7] If Lora fulfilled her desire, she would then subsequently pose no threat to others, then Lora should not spend time in isolation, and she would not need rehabilitation. Lora may then reasonably expect to spend, as a consequence of her action, a few days in a rehabilitative center, fully incapacitated, and have some tests run on her in the meantime, to determine whether she is dangerous or not. However, a few days or even a month of isolation is nothing compared to Lora's goal – killing her mother. After giving some thoughts to this scenario, Lora murders Klara. If Lora's calculations are right, and she does not pose any threats to society, then Lora should be set free, in accordance with the quarantine model.

Pereboom (2001, 175), considered a case, where people kill their spouses out of anger and some people may think that such a person is not dangerous for others. But Pereboom argued that those who are disposed to kill once are likely to be often disposed to be criminally violent. That would justify their detention. However, whether offenders are dangerous cannot be established by simply appealing to mere intuitions or generalizations. The probability of reoffending may turn out to be very different for different spouses. The risk of reoffending would, thus, depend on many different factors including, but not limited to, alcohol and drug addiction, history of

---

[7] There is no contradiction with what we said in the previous section. In fact, there are certainly cases, both positive and negative, in which the evaluation of the danger is pretty safe. Lora's case is a case in point.

violent behavior, criminal record and lack of employment, personal traits, situational analysis of the murder, etc.

For some of the murderous spouses, the chance of recidivism would be high and for others it may be very low. In Lora's case, in which we just described her risk of reoffending would most likely be negligible (Lora wouldn't have a reason to kill someone else). Accurate risk-assessment procedures would establish that offenders like Lora are not dangerous.[8] These people would therefore reasonably expect that they have a free pass for a one-time offense. That sort of knowledge may well serve as an incentive for committing a crime. That very possibility, we believe, poses a critical problem for the quarantine model. How many people would there be out there who would just want to kill one and only one very annoying person (e.g., a celebrity, a politician, or a relative)?

Here we want to be very specific. We did not introduce Lora's case to second vengeful intuitions against some criminals, who can be potentially released unpunished. The quarantine model does not aim to punish. Pereboom and Caruso easily counter such intuitions by pointing out that they bring more harm than good, and we should rather get rid of them. Instead, we showed that *the quarantine model could encourage Lora to murder her mother*. The current mixed criminal system does not risk encouraging such crimes because it has strong deterrence mechanisms against them. Today Lora would simply know that she would go to prison if she got caught. This is a clear deterrent. Yet, for the quarantine model, Lora poses a significant problem.

Caruso and Pereboom may argue that their quarantine model is more humane than retributivism, especially given the assumption that Lora is not morally responsible in the basic desert sense, for having killed her mother. The quarantine model would allow Lora to avoid needless suffering in prison. There wouldn't be a reason to keep her in prison if she does not pose any threat to society. The central problem, in this case, is not that the quarantine model fails to ratify our intuitive verdict that Lora should be punished; rather, we believe that the problem is that the quarantine model seems to facilitate and perhaps even encourage such crimes.

In response to Smiliansky's like cases, Pereboom incorporated a general deterrence component into his model. Pereboom (2017) writes: "The model I advocate includes general deterrence by monetary penalties and short-term prison sentences. This yields a response to one example Smilansky provides".

However, Lora's case, which is not unrealistic, shows how ineffective the quarantine model would be in dealing with it. Crucially, there could be many other examples of one-time offenders that will raise the same issue. For example, euthanasia, "motivated by compassion" for terminally ill patients, or patients with severe dementia could greatly increase in societies with a high percentage of elderly people. All those "killers" of elderly relatives would be cognizant that they are not poised to be sanctioned, since they are supposed to act out of compassion and are

---

[8] Neither Pereboom nor Caruso think it is justifiable to quarantine a person without a "triggering" event (an offense or some clear warning signs), which is necessary to begin an analysis of dangerousness (Caruso 2021; Pereboom 2001, 2014). Authorities would not estimate Lora's risk score before the murder, and the official estimate is not something Lora and other potential offenders would have at hand. However, an educated person should be able to understand what her "dangerousness estimate" would be like after a certain offense given the factors that crime prevention units consider.

not dangerous people overall. Nevertheless, it seems reasonable to deter such behavior, as euthanasia—when legally permitted- should be regulated by the state, and only administered by medical professionals.

Moreover, the effectiveness of monetary penalties and short-term prison sentences, as deterrence measures, is questionable. There is not enough empirical data about the effectiveness of such measures against serious crimes (more on it below). Pereboom, Caruso and their critics can only speculate on this point. Probably fines will be a very weak deterrent for the sort of crimes (non-financially motivated) that we described in our example. In addition, fines would likely be ineffective as a deterrent for extremely rich people, who could afford to pay virtually any fine (even if such fines would be proportional to their income, to be charitable with one of Pereboom's suggestions and as is done in some countries). Pereboom acknowledges the problem with fines and suggests the implementation of short prison sentences as extrema ratio (Pereboom, 2018). Short-term prison sentences supposedly should serve as a stronger deterrence without hindering the prospects for a life lived at a reasonable level of flourishing. Pereboom refers here to Mark Kleiman, who argues that short prison sentences are often effective deterrents, especially in combination with a high expectation of being apprehended (Kleiman, 2009).

However, it is important to distinguish between the idea of a short-term prison sentence and the idea of reducing the longevity of a sentence in general. Short-term prison sentences are typically sentences that require the inmate to serve less than 12 months. One thing is to deter thieves with a three-months sentence, quite another is to reduce the term of a sentence for murder from fifteen years to, say, five years. Shorter than average prison sentences may very well serve as a deterrence and be more humane, but this is not the same thing as short-term prison sentences. We do not intend to enter, here, the empirical debate over the effectiveness of short-term prison sentences, as we are not competent enough, but two points need to be made.

Firstly, a methodological point: all the available data on short-term prison sentences is obtained on relatively minor crimes. Secondly, the effectiveness of short-term prison sentences is disputed, even if applied to minor crimes (Trebilcock, 2011). Thus, short-term prison sentences may not be enough to deter potential one-time offenders, unless one is prepared to endorse and promote a hybridized model, one that includes as key traits, some of the features of general deterrence models (such as classical detention in prison), which are not justified based on retributivism, but are necessarily very similar with those found in many of our current criminal systems.

In addition, even if monetary penalties and short-term prison sentences were to be an efficient deterrent (which is arguable, as we have shown above), there would also be a conceptual difficulty in implementing fines and sentences into a quarantine model. Quarantine regulations do not justify, for the sake of general deterrence, the application of sanctions for being infected or being exposed to infected individuals. There is simply no need to punish people for deterrence purposes, for contracting a disease or for being exposed to it, as an illness is an undesirable consequence already, in most cases. Caruso (2021) recognizes that allowing such measures for the purpose of general deterrence goes beyond what is permitted by the quarantine model. Specifically, Caruso argues that it moves the theory closer to consequentialist

models of punishment and it introduces a punitive component in their model that he wants to avoid completely (2021, 312–313).

Thus, if Pereboom wants to keep fines and short-term prison sentences for deterrence, he should concede that the quarantine analogy, on which his model is grounded, is no longer playing a central role in his account. Pereboom may reply by saying incarceration and monetary fines are reserved for those who violate quarantine. However, those people are not punished because they have been exposed to a contagious disease. Punishment is inflicted upon them for the intentional violation of enforced quarantine and for bringing danger upon the public. In Lora's case and analogous ones involving one-time offenders, wrongdoers should not be quarantined for preventive reasons, as they pose no genuine threat to society beyond the crime they committed.

Caruso (2021, 316) thinks that the quarantine model already has the resources to deal with low-level, nonviolent crime. Caruso proposes that various liberty-limiting measures short of incapacitation could suffice to address low-level crimes (such as speeding or shoplifting). However, this leaves the question about deterrence of non-dangerous offenders unanswered. Caruso himself sets a very high epistemic bar for determining the dangerousness of a person and he argues that only those who pose a serious threat to society can be incapacitated (2021, 320–322). Given the circumstances of the case study we discussed above (involving Lora and Klara) any judge would struggle to conclude -beyond reasonable doubt- that Lora poses a serious threat to society. In other words, it would seem as if society and other people would not need to protect themselves from her.

So, for Caruso it seems that the best option would be to bite the bullet and agree that Lora should be set free. He already concedes that if that violent offender is successfully rehabilitated within two years, then we should set him free in accordance with the quarantine model (2021, 277). The scenario with effective rehabilitation was fictional (Shariff et al., 2014). The main difference from our scenario is that Lora does not need rehabilitation and she is not dangerous right after her crime. And Caruso is quite clear on who shouldn't be incapacitated in this model: '(a) individuals who are not a serious threat to society should not be incapacitated, (b) no one should be incapacitated longer than is absolutely necessary (where this is determined by the continued threat the individual poses to society)' (2021, 278). Lora poses no serious threat to society and her incapacitation is not necessary. Therefore, according to Caruso, we have no reasons to incarcerate her.

Pereboom and Caruso may insist that such cases are exceptional cases, or mere products of philosophical imagination. However, they are not. Cases when a person commits a serious crime without posing reasonable risk of reoffending do exist in the real world, think about the Albanian blood feud (Gjakmarrja)[9] or about the case of Vitaly Kaloyev[10] (both reviewed and discussed in Levin et al., 2021).

---

[9] Gjakmarrja refers to a traditional Albanian practice of seeking vengeance. If a member of a family is killed, the code may require that the responsible party or a male member of their family be killed in retribution.

[10] Vitaly Kaloyev was a Russian architect who murdered the air traffic controller Peter Nielsen in 2004. Kaloyev believed Nielsen was responsible for a 2002 mid-air collision over the skies of Germany that killed his family.

Furthermore, a demand only for real-life scenarios would be quite inconsistent from proponents of the quarantine model, who themselves often appeal to quite extravagant thought experiments to justify their intuitions (e.g., Pereboom, 2001, 2014). The existence of real-world cases and the logical possibility of serious offenders who do not constitute a danger to society forces Pereboom to resort to standard consequentialist reasoning in order to close a gap in his model. Caruso acknowledges that such a move is against the spirit of the quarantine model (2021). However, he does not provide a solution to the problem.

5 Conclusion

The quarantine model remains one of the cornerstones of contemporary debates on criminal justice. In this article we raised two main criticisms against it. These are related to: (a) how the quarantine model lacks a reliable way of determining who is dangerous and, therefore, the need to heavily rely on risk-assessment tools, which is -as we have shown- morally problematic; (b) the possibility that the quarantine model can encourage certain crimes. A possible way out, which, however, authors like Pereboom and Caruso do not want to choose, is to embrace a fully-fledged version of consequentialism, which is not interested in the rights of, or potential harm to, a minority of offenders, so long as the greater welfare of the majority is attained. Perhaps Pereboom and Caruso could suggest a convincing reply to our arguments within the theoretical framework of the quarantine model. We indeed share with them their aspiration to improve the functioning of current criminal justice systems. However, our critiques -if not unfounded- should contribute to making the quarantine model more coherent and humane.

Declarations

Ethics approval　n/a

Consent to participate　n/a

Consent for publication　n/a

Conflicts of Interest/Competing Interests　n/a

References

Aboodi, R., Borer, A., & Enoch, D. (2008). Deontology, individualism, and uncertainty: A reply to Jackson and Smith. *The Journal of Philosophy, 105*(5), 259–272.

Ægisdóttir, S., White, M. J., Spengler, et al. (2006). The meta-analysis of clinical Judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In K. Martin (Ed.), *Ethics of data and analytics* (pp. 254–264). Auerbach Publications.

Bedau, H. A., & Kelly, E. (2019). Punishment. *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.) https://plato.stanford.edu/archives/win2019/entries/punishment/

Berk, R., Heidari, H., Jabbari, S., et al. (2018). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research, 50*(19), 3–44.

Berman, M. N. (2008). Punishment and justification. *Ethics, 118*(2), 258–290.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*(6), 567–590.

Caruso, G. D. (2016). Free will skepticism and criminal behavior: A public health-quarantine model. *Southwest Philosophy Review, 32*(1), 25–48. https://doi.org/10.5840/swphilreview20163214

Caruso, G. D. (2017). *Public health and safety: The social determinants of health and criminal behavior*. ResearchersLinks Books.

Caruso, G. D. (2020). Justice without retribution: An epistemic argument against retributive criminal punishment. *Neuroethics, 13*(1), 13–28.

Caruso, G. D. (2021). *Rejecting retributivism: Free will, punishment, and criminal justice*. Cambridge University Press.

Caruso, G. D., & Dennett, D. C. (2021). *Just deserts: debating free will*. Polity Press.

Cashmore, A. R. (2010). The Lucretian Swerve: The Biological basis of human behavior and the criminal justice system. *Proceedings of the National Academy of Sciences of the United States of America, 107*(10), 4499–4504.

Corrado, M. L. (2016). Chapter one. Two models of criminal justice. *UNC Legal Studies Research Paper No. 2757078*. https://doi.org/10.2139/ssrn.2757078

Corrado, M. L. (2019). Criminal quarantine and the burden of proof. *Philosophia, 47*(4), 1095–1110.

Crawford, K. (2021). *The atlas of AI: Power: Politics, and the planetary costs of artificial intelligence*. Yale University Press.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*, eaao5580.

Gillies, D. (2018). *Causality, probability, and medicine*. Routledge.

Gordis, L. (2014). *Epidemiology* (5th ed.). Elsevier/Saunders.

Hamilton, M. (2015). *Risk-needs assessment: Constitutional and ethical challenges* (SSRN Scholarly Paper No. ID 2506397). Social Science Research Network.

Harcourt, B. E. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter, 27*, 237–243.

Harris, S. (2012). *Free will*. Free Press.

Haarsma, G., Davenport, S., White, D. C., et al. (2020). Assessing risk among correctional community probation populations: Predicting reoffense with mobile neurocognitive assessment software. *Frontiers in Psychology, 10*, 2926.

Henry, T. K. S. (2021). Revolving doors: Examining the effect of race and ethnicity on discretionary decision-making in parole revocations. *Am J Crim Just, 46*, 279–297.

Huemer, M. (2010). Lexical priority and the problem of risk: Lexical priority and the problem of risk. *Pacific Philosophical Quarterly, 91*(3), 332–351. https://doi.org/10.1111/j.1468-0114.2010.01370.x

Husak, D. (2013). Retributivism in extremis. *Law and Philosophy, 32*(1), 3–31.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.

Kershnar, S. (2000). A defense of retributivism. *International Journal of Applied Philosophy, 14*(1), 97–117.

Kleiman, M. A. R. (2009). *When brute force fails*. Princeton University Press.

Kolber, A. (2018). Punishment and moral risk. *University of Illinois Law Review, 2*, 487–532.

Lemos, J. (2016). Moral concerns about responsibility denial and the quarantine of violent criminals. *Law and Philosophy, 35*(5), 461–483.

Lemos, J. (2018). *A pragmatic approach to libertarian free will*. Routledge.

Lemos, J. (2019). A moral/pragmatic defense of just deserts responsibility. *Journal of Information Ethics, 28*(19), 73–94.

Laqueur, H., & Copus, R. (2022). *An algorithmic assessment of parole decisions* (SSRN Scholarly Paper No. 4078152).

Levin, S., Farina, M., & Lavazza, A. (2021). Dealing with criminal behavior: The inaccuracy of the quarantine analogy. *Criminal Law and Philosophy*. https://doi.org/10.1007/s11572-021-09608-2

Levy, N. (2011a). Luck and free will. In *New waves in philosophy of action* (pp. 173–188). Palgrave Macmillan.

Levy, N. (2011b). *Hard luck: How luck undermines free wil and moral responsibility*. Oxford University Press.

Libet, B., Wright, E. W., Jr., & Gleason, C. A. (1983). Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalography and Clinical Neurophysiology, 56*(4), 367–372.

Meyerson, D. (2009). Risks, rights, statistics, and compulsory measures. *Sydney Law Review, 31*, 507–535.

Menninger, K. A. (1968). *The crime of punishment*. Penguin Books.

Moore, M. (1997). *Placing blame: A theory of the criminal law*. Oxford University Press.

Murphy, J. G. (1979). Cruel and unusual punishments. In M. A. Stewart (Ed.), *Law, morality and rights* (pp. 373–404). Springer, Netherlands.

Murphy, J. G. (2007). Legal moralism and retribution revisited. *Criminal Law and Philosophy, 1*(1), 5–20. https://doi.org/10.1007/s11572-006-9000-3

Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.

Pereboom, D. (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.

Pereboom, D. (2018). Incapacitation, reintegration, and limited general deterrence. *Neuroethics*. https://doi.org/10.1007/s12152-018-9382-7

Pereboom, D. (2019). Free will skepticism and prevention of crime. In E. Shaw, D. Pereboom, & G. D. Caruso (Eds.), *Free will skepticism in law and society.* Cambridge University Press.

Pereboom, D. (2020). Incapacitation, reintegration, and limited general deterrence. *Neuroethics, 31*(1), 87–97.

Pereboom, D., & Caruso, G. D. (2018). Hard-incompatibilist existentialism: neuroscience, punishment, and meaning in life. In Gregg D. Caruso & Owen Flanagan (Eds.), *Neuroexistentialism.* Oxford University Press.

Rawls, J. (1955). Two concepts of rules. *The Philosophical Review, 64*, 3–32.

Raaijmakers, S. (2019). Artificial intelligence for law enforcement: Challenges and opportunities. *IEEE Security & Privacy, 17*(5), 74–77.

Reiling, A. D. (2020). Courts and artificial intelligence. *International Journal for Court Administration, 11*(2), 8.

Russell, P. (2021). Hume on Free Will. In Zalta E. N. (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2021/entries/hume-freewill/

Shafer-Landau, R. (1996). The failure of retributivism. *Philosophical Studies, 82*(3), 289–316.

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science, 25*(8), 1563–1570. https://doi.org/10.1177/0956797614534693

Schoeman, F. D. (1979). On incapacitating the dangerous. *American Philosophical Quarterly, 16*, 27–35.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., et al. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health, 13*(3), 193–206.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology, 54*, 680–712.

Slobogin, C. (2021). Preventive justice: How algorithms, parole boards, and limiting retributivism could end mass incarceration. *Wake Forest Law Review, 56*(1), 97–168.

Smilansky, S. (2017). Pereboom on punishment - funishment, innocence, motivation, and other difficulties. *Criminal Law and Philosophy, 11*(3), 591–603.

Smilansky, S. (2019). Free will denial and deontological constraints. In E. Shaw, D. Pereboom, & G. D. Caruso (Eds.), *Free will skepticism in law and society: Challenging retributive justice* (pp. 29–42). Cambridge University Press.

Sox, H. C., Jr. (1986). Diagnostic decision: Probability theory in the use of diagnostic tests: An introduction to critical study of the literature. *Annals of Internal Medicine, 104*(1), 60–66.

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review, 66*, 803–872.

Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 75*(1/2), 5–24.

Tadros, V. (2017). Doing without desert. *Criminal Law and Philosophy, 11*(3), 605–616.

Trebilcock, J. (2011). *No winners: The reality of short term prison sentences*. The Howard League for Penal Reform.

Vilhauer, B. (2013). Persons, punishment, and free will skepticism. *Philosophical Studies, 162*(2), 143–163.

Von Hirsch, A. (1992). Proportionality in the philosophy of punishment. *Crime and Justice, 16*, 55–98.

Walen, A. (2020). Retributive justice. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2020/entries/justice-retributive/

Waller, B. N. (2011). *Against moral responsibility*. MIT Press.

Waller, B. N. (2015). *Restorative free will: Back to the biological base*. Lexington Books.

Ward, T., Melser, J., & Yates, P. M. (2007). Reconstructing the risk-need-responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior, 12*, 208–228.

Wegner, D. M. (2002). *The illusion of conscious will. A Bradford Book*. The MIT Press.

Xu, Z. (2022). Human judges in the era of artificial intelligence: Challenges and opportunities. *Applied Artificial Intelligence, 36*(1), 2013652.