

A Non-Punitive Alternative to Retributive Punishment

Gregg D. Caruso and Derk Pereboom

Penultimate Draft for:

Routledge Handbook of the Philosophy and Science of Punishment

Edited by Farah Focquaert, Bruce Waller, and Elizabeth Shaw

Punishment is the intentional imposition of a penalty upon a person or group of persons for conduct that is represented, either truly or falsely, as immoral or in violation of a law. Everyday examples include poker players excluding someone who has cheated from further play, and a teacher giving a group of pupils a “time out” for disrespectful behavior. *Legal punishment* is a specific sort of punishment; it is the intentional imposition of a penalty for conduct that is represented, either truly or falsely, as a violation of a law of the state, where the imposition of that penalty is sanctioned by the state’s authority (Zimmerman 2011; Boonin 2008; Bedau and Kelly 2015).

One of the most prominent justifications of legal punishment, historically and currently, is *retributivism*, according to which wrongdoers *deserve* the imposition of a penalty solely for the backward-looking reason that they have knowingly done wrong. Michael S. Moore, a leading retributivist, highlights this purely backward-looking nature of retributivist justification:

[R]etributivism is the view that we ought to punish offenders because, and only because, they deserve to be punished. Punishment is justified, for a retributivist, solely by the fact that those receiving it deserve it. Punishment may deter future crime, incapacitate dangerous persons, educate citizens in the behaviour required for a civilized society, reinforce social cohesion, prevent vigilante behaviour, make victims of crime feel better, or satisfy the vengeful desires of citizens who are not themselves crime victims. Yet for

the retributivist these are a happy surplus that punishment produces and form no part of what makes punishment just: for a retributivist, deserving offenders should be punished even if the punishment produces none of these other, surplus goof effects.” (2010: 153; see also 1987, 1993)

Retributivism is also endorsed, for example, by Stephen Kershnar (2000, 2001), Douglas Husak (2000), and arguably by Immanuel Kant (1790). It’s important to emphasize that the desert invoked in retributivism (in the classical or strict sense) is *basic* in the sense that it is not in turn grounded in forward-looking reasons such as securing the safety of society or the moral improvement of criminals, or, more broadly, in the value of consequences. Thus, for the retributivist, the claim that persons are morally responsible for their actions in the *basic desert* sense is crucial to the state’s justification for giving them their *just deserts* in the form of punishment for violations of the state’s laws. Retributivists typically also hold, in addition, that just punishments must be *proportional to wrongdoing*. Both the justificatory thesis and the proportionality requirement for punishments are reflected in Mitchell Berman’s statement of retributivism: “A person who unjustifiably and inexcusably causes or risks harm to others or to significant social interests deserves to suffer for that choice, and he deserves to suffer in proportion to the extent to which his regard or concern for others falls short of what is properly demanded of him” (2008: 269).

In the American criminal legal system, the retributivist justification of legal punishment and the attendant proportionality requirement are widely embraced. A number of sentencing guidelines in the U.S. have adopted the retributivist conception of desert as their core principle,¹ and it is increasingly given deference in the “purposes” section of state criminal codes,² where it can be the guiding principle in the interpretation and application of the code’s provisions.³

Indeed, the American Law Institute recently revised the Model Penal Code so as to set desert as the officially dominant principle for sentencing.⁴ American courts have identified desert as the guiding principle in a variety of contexts,⁵ as with the Supreme Court's enthrone retributivism as the "primary justification for the death penalty"⁶ (Robinson 2008: 145-146). But despite its prevalence in the criminal law, an important philosophical question remains: Does retributivism as a theory of legal punishment withstand scrutiny? In particular, is state-sanctioned intentional imposition of penalties for violating the state's laws ever basically deserved?

Free Will Skepticism

One reason to think that no human agents ever basically deserve legal punishment, as retributivism presupposes, is that we lack the control in action, that is, the free will, required for moral responsibility in the *basic desert* sense, which we define as follows:

For an agent to be morally responsible for an action in the basic desert *sense* is for it to belong to her in such a way that she would deserve blame if she understood that it was morally wrong, and she would deserve credit or perhaps praise if she understood that it was morally exemplary. The desert invoked here is *basic* in the sense that the agent, to be morally responsible, would deserve the blame or credit just because she has performed the action, given sensitivity to its moral status, and not by virtue of consequentialist or contractualist considerations. (Pereboom 2001, 2014; cf. Feinberg 1970)

Free will skeptics maintain that since what we do and the way we are is ultimately the result of factors beyond our control, we are never morally responsible for our actions in the basic desert sense (see, e.g., Pereboom 2001, 2014; Levy 2011; Caruso 2012). Free will skepticism accordingly presents a serious challenge to retributivism: if we never basically deserve blame

just because we have knowingly done wrong, neither do we ever basically deserve punishment just because we have knowingly done wrong.

Free will skepticism, one might object, allows for non-basically deserved blaming and praising – for example, blaming that invokes desert grounded in consequentialist (e.g., Daniel Dennett 1984; Manuel Vargas 2013) or contractualist (e.g., Ben Vilhauer 2013) considerations. On one type of revisionary account, our practice of holding agents morally responsible in a desert sense should be retained, not because we are in fact morally responsible in this sense, but because doing so would have the best consequences relative to alternative practices. Daniel Dennett (1984) advocates a version of this position, as does Manuel Vargas (2013). But punishment justified in this way would not be genuinely retributivist, since its ultimate justification would be consequentialist, and this is incompatible with retributivism as it has traditionally been understood.

The case we favor for free will skepticism features distinct arguments that target three rival views, *event-causal libertarianism*, *agent-causal libertarianism*, and *compatibilism*, and then claims that the skeptical position is the only defensible position that remains standing. We begin by considering the libertarian positions. According to event-causal libertarianism, actions are caused solely by way of events, standardly conceived as objects having properties at times, and some type of indeterminacy in the production of actions by appropriate events is held to be a decisive requirement for moral responsibility (Balaguer 2009; Ekstrom 2000; Kane 1996).

According to agent-causal libertarianism, free will of the sort required for moral responsibility is accounted for by the existence of agents who possess a causal power to make choices without being causally determined to do so (Chisholm 1976; Clarke 2003; Griffith 2010; Kant 1781/1787/1987; O'Connor 2000; Reid 1788/1983; Taylor 1974). In this view, it is

essential that the causation involved in an agent's making a free choice is not reducible to causation among events involving the agent, but is rather irreducibly an instance of the agent-as-substance causing a choice not by way of events. The agent, fundamentally as a substance, has the causal power to cause choices without being determined to do so.

Here, in brief, is the reason we reject event-causal libertarianism. For an agent to be morally responsible for a decision in the basic desert sense, she must exercise a certain type and degree of control in making that decision. In the event-causal libertarian picture, only events are causes, and free decisions are causally undetermined. The causally relevant events antecedent to a decision—most prominently agent-involving events—accordingly leave it open whether the decision occurs, and thus do not settle whether it occurs. Settling whether a decision occurs is a kind of control in action, and the event-causal libertarian is a causalist about control, specifying that control in action is a causal matter. But because on this view all causation is event-causation, the agent can have no role in settling whether a decision occurs beyond the role it plays in agent-involving events. Hence, given the indeterminism required for a free decision, the agent cannot settle whether the decision occurs. Therefore, in the event-causal libertarian picture an agent cannot have the control required for being morally responsible in the basic desert sense for a decision. Since the agent “disappears” at the crucial juncture in the production of the decision – when its occurrence is to be settled – Pereboom calls this the *disappearing agent argument* (Pereboom 2004, 2014, 2017a).

The agent-causal libertarian's response to this difficulty is to specify a way in which the agent might possess the required power to settle in indeterministic contexts. The solution on offer is to reintroduce the agent as a cause, this time not merely as involved in events, but rather fundamentally as a substance. The agent-causal libertarian proposes that we possess a distinctive

causal power—a power for an agent, fundamentally as a substance, to cause a decision without being causally determined to do so, and thereby to settle whether it occurs (Clarke 2003; Chisholm 1976; Griffith 2010; Kant 1781/1787/1987; O'Connor 2000).

We don't think that the agent-causal libertarian picture is ruled out as incoherent. But it does face difficulties that undermine its credibility as a ground for a retributivist justification of punishment, in particular given the epistemic argument that we set out below. Immanuel Kant (1781/1787/1987) expresses a significant concern for this view, that it might not be reconcilable with what we would expect given our best empirical theories. He held that the physical world, as part of the world of appearance, is governed by deterministic laws, whereas the “transcendentally free” agent-cause would exist not as an appearance, but as a thing in itself. In his agent-causal picture, when an agent makes a free decision, she causes the decision without being causally determined to do so. On the route from this undetermined decision to its effects, changes in the physical world, for example, in the agent's brain or some other part of her body, are produced. However, it would seem that we would at these points encounter divergences from the deterministic laws, since the changes in the physical world that result from the undetermined decision would themselves not be causally determined. One might respond that it is possible that the physical alterations that result from free decisions just happen to dovetail with what could in principle be predicted on the basis of the deterministic laws, so nothing actually occurs that diverges from them. But this reconciliationist proposal would appear to involve coincidences too wild to be credible. Consequently, agent-causal libertarianism appears not to be reconcilable with a deterministic law-governed physical world. Pereboom (1995, 2001, 2014) also argues that similar sorts of wild coincidences would plague a view according to which agent-caused free choices would reconcile with indeterministic and probabilistic laws. If the libertarian agreed to

these arguments from wild coincidences, she might propose that there are indeed departures from the probabilities that we would expect given the physical laws, thus abandoning the reconciliationist project. Roderick Chisholm (1964) proposes a solution of this sort. An objection to this proposal is that we would seem to lack any evidence that departures from the known physical laws occur in the brain when we make decisions.

The remaining and most popular alternative to skepticism about free will is compatibilism. One way to argue against compatibilism begins with the intuition that if an agent is causally determined to act by, for example, scientists who unbeknownst to her manipulate her brain, then she is not morally responsible for that action, even if she satisfies the prominent compatibilist conditions on moral responsibility (Ginet 1990; Kane 1996; Mele 1995, 2006; Pereboom 1995, 2001, 2014; Taylor 1974; van Inwagen 1983). The key subsequent step is that there are no differences between such manipulated agents and their ordinary deterministic counterparts that can justify the claim that the manipulated agents are not morally responsible while the determined agents are.

Pereboom's multiple-case version of such an argument first of all develops examples of an action that results from an appropriate sort of manipulation and in which the prominent compatibilist conditions on moral responsibility are satisfied (Pereboom 1995, 2001, 2014). In the setup, in each of the four cases the agent commits a crime, say homicide, for the sake of some personal advantage. The cases are designed so that the crime conforms to the prominent compatibilist conditions; for instance, the action meets the reasons-responsiveness condition proposed by John Fischer and Mark Ravizza (1998): his desires can be modified by, and some of them arise from, his rational consideration of the reasons, and if he understood that the bad consequences for himself that would result from the crime would be more severe than they are

actually likely to be, he would have refrained for that reason. The individual manipulation cases featured in the argument, considered separately, indicate that it is possible for an agent not to be morally responsible in the basic desert sense even if these prominent compatibilist conditions are satisfied, and that these conditions are therefore insufficient. The argument acquires additional force by virtue of setting out three manipulation cases, each of which is progressively more like the fourth case, in which the action is causally determined in a natural way. The first case involves manipulation that is immediately causally determining – the neuroscientists directly control the agent’s reasoning from moment to moment. The second is just like the first, except that it restricts the manipulation to a time at the beginning of the agent’s life. The third case is also similar, except that the manipulation results from the strict upbringing on the part of his community. The fourth is the natural or ordinary deterministic case. The aim is to formulate these cases so that it is impossible to point to a difference between any two adjacent cases that would explain in a principled way why the agent would not be morally responsible in the basic desert sense in one but would be in the other. The conclusion is that the agent is not morally responsible in this sense in all four cases, and that the best explanation for this must be that he is causally determined by factors beyond his control in each of them. This result conflicts with the compatibilist’s central claim.

Further Reasons to Reject Retributivism

Even if one is not convinced by the arguments against the sort of free will required for basic desert, there remains a strong *epistemic argument* against causing harm on retributivist grounds that is sufficient for the rejection of retributive legal punishment (see Caruso 2018). This is because the burden of proof lies on those who want to intentionally inflict harm on others to provide good justification for such harm. This means that retributivists who want to justify legal

punishment on the assumption that agents are free and moral responsible (and hence *justly deserve* to suffer for the wrongs they have done) must justify that assumption in a way that meets a high epistemic standard of proof since the harms caused in the case of legal punishment are often quite severe. It is not enough to simply point to the mere possibility that agents possess libertarian or compatibilist free will. Nor is it enough to say that the skeptical arguments against free will and basic desert moral responsibility fail to be conclusive. Rather, a positive and convincing case must be made that agents are in fact morally responsible in the basic desert sense, since it is the backward-looking desert of agents that retributivists take to justify the harm caused by legal punishment.

The epistemic argument against retributive legal punishment can be summarized as follows (Caruso 2018):

- (1) Legal punishment inflicts harms on individuals and the justification for such harms must meet a high epistemic standard. If it is significantly probable that one's justification for harming another is unsound, then, *prima facie*, that behavior is seriously wrong.
- (2) 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 (appropriately qualified and under the constraint of proportionality).
- (3) If the justification for the assumption that agents are morally responsible in the basic desert sense and hence justly deserve to suffer for the wrongs they have done does not meet the high epistemic standard specified in (1), then retributive legal punishment is *prima facie* seriously wrong.

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

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

This argument builds on previous work done by Double (2002), Vilhauer (2015), and Pereboom (2001, 2014), and has been more recently developed and defended by Caruso (2018), Corrado (2017), and Shaw (2014). We take our previous statements of the argument to be sufficient and have nothing new to add here other than to reiterate that all extant accounts of basic desert moral responsibility still fail to satisfy the high epistemic standard needed to justify retributive legal punishment.

Several additional considerations also count against retributivism. Retributivist sentiments may be grounded in vengeful desires, and that therefore retribution has little more plausibility than vengeance as a morally sound policy for action. Acting on vengeful desires might be wrong for the following sort of reason. Although acting on such desires can bring about pleasure or satisfaction, no more of a moral case can be made for acting on them than can be made for acting on sadistic desires, for example. Acting on sadistic desires can bring about pleasure, but in both cases acting on the desire aims at the harm of the one to whom the action is directed, and in neither case does acting on the desire essentially aim at any good other than the pleasure of its satisfaction. But then, since retributivist motivations are disguised vengeful desires, acting for the sake of retribution is also morally wrong (Pereboom 2001, 2014).

Moreover, supposing that the requisite capacity for control is in place, and that basic desert could be secured as good or right, we can ask whether the state has the right to invoke it in

justifying punishment. The legitimate functions of the state are generally held to include protecting its citizens from significant harm, and providing a framework for human interaction to proceed without significant impairment. These roles arguably underwrite justification that in the first instance appeals to prevention of crime. But these roles have no immediate connection to the aim of apportioning punishment in accord with basic desert. The concern can be made vivid by considering the proposal that the state set up a well-funded set of institutions designed to comprehensively and fairly distribute rewards on the grounds of basic desert (Pereboom 2014).

For these and other reasons (see, e.g., Boonin 2008; Zimmerman 2011; Caruso 2017, 2019), we conclude that retributivism as a justification of legal punishment should be rejected. We now turn to our non-retributive alternative for addressing criminal behavior—the *public health-quarantine model*.

The Public Health-Quarantine Model

The public health-quarantine model is based on an analogy with quarantine and draws on a comparison between treatment of dangerous criminals and treatment of carriers of dangerous diseases (see Pereboom 2001, 2013, 2014; Caruso 2016, 2017; Pereboom and Caruso 2018). In its simplest form, it can be stated as follows: (1) Free will skepticism maintains that criminals are not morally responsible for their actions in the basic desert sense; (2) plainly, many carriers of dangerous diseases are not responsible in this or in any other sense for having contracted these diseases; (3) yet, we generally agree that it is sometimes permissible to quarantine them, and the justification for doing so is the right to self-protection and the prevention of harm to others; (4) for similar reasons, even if a dangerous criminal is not morally responsible for his crimes in the basic desert sense (perhaps because no one is ever in this way morally responsible) it could be *as*

legitimate to preventatively detain him as to quarantine the non-responsible carrier of a serious communicable disease.

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, we want to resist this strategy (see Pereboom and Caruso 2018). Instead, our view maintains that incapacitation of the seriously dangerous is justified on the ground of the right to harm in self defense and defense of others. That we have this right has broad appeal, much broader than utilitarianism or consequentialism has. In addition, this makes the view more resilient to a number of objections and provides a more resilient proposal for justifying criminal sanctions than other non-retributive options (see, e.g., Pereboom 2017b; Pereboom and Caruso 2018; Caruso 2019). One advantage it has, say, over consequentialist deterrence theories is that it has more restrictions placed on it with regard to using people merely as a means. For instance, as it is illegitimate to treat carriers of a disease more harmfully than is necessary to neutralize the danger they pose, treating those with violent criminal tendencies more harshly than is required to protect society will be illegitimate as well. In fact, in all our writings on the subject, we have always maintained the *principle of least infringement*, which holds that the least restrictive measures should be taken to protect public health and safety (Caruso 2016, 2017; Pereboom and Caruso 2018). This ensures that criminal sanctions will be proportionate to the danger posed by an individual, and any sanctions that exceed this upper bound will be unjustified.

Second, the quarantine model places several constraints on the treatment of criminals (see Pereboom 2001, 2014). First, as less dangerous diseases justify only preventative measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate

restraints (Pereboom 2014: 156). We do not, for instance, quarantine people for the common cold even though it has the potential to cause you some harm. Rather, we restrict the use of quarantine to a narrowly prescribed set of cases. Analogously, on our model the use of incapacitation should be limited to only those cases where offenders are a serious threat to public safety and no less restrictive measures were available. In fact, for certain minor crimes perhaps only some degree of monitoring could be defended. Secondly, the incapacitation account that results from this analogy demands a degree of concern for the rehabilitation and well-being of the criminal that would alter much of current practice. Just as fairness recommends that we seek to cure the diseased we quarantine, so fairness would counsel that we attempt to rehabilitate the criminals we detain (Pereboom 2014: 156). Rehabilitation and reintegration would therefore replace punishment as the focus of the criminal justice system. Lastly, if a criminal cannot be rehabilitated and our safety requires his indefinite confinement, this account provides no justification for making his life more miserable than would be required to guard against the danger he poses (Pereboom 2014: 156).

In addition to these restrictions on harsh and unnecessary treatment, the model also advocates for a broader approach to criminal behavior that moves beyond the narrow focus on sanctions. As Caruso (2016, 2017, 2019) has proposed, and Pereboom has endorsed (Pereboom and Caruso 2018), we recommend placing the quarantine analogy within the broader justificatory framework of *public health ethics*. Public health ethics not only justifies quarantining carriers of infectious diseases on the grounds that it is necessary to protect public health, it also requires that we take active steps to *prevent* such outbreaks from occurring in the first place. Quarantine is only needed when the public health system fails in its primary function. Since no system is perfect, quarantine will likely be needed for the foreseeable future, but it should *not* be the

primary means of dealing with public health. The analogous claim holds for incapacitation.

Taking a public health approach to criminal behavior would allow us to justify the incapacitation of dangerous criminals when needed, but it would also make prevention a *primary function* of the criminal justice system. So instead of myopically focusing on punishment, the public health-quarantine model shifts the focus to identifying and addressing the systemic causes of crime, such as poverty, low social economic status, systematic disadvantage, mental illness, homelessness, educational inequity, exposure to abuse and violence, poor environmental health, and addiction (see Caruso 2017).

We maintain that since the *social determinants of health* and the *social determinants of criminal behavior* are broadly similar (Caruso 2017), 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 would require investigating how social inequities and systemic injustices affect health outcomes and criminal behavior, how poverty affects brain development, how offenders often have pre-existing medical conditions (especially mental health issues), how homelessness and education affects health and safety outcomes, how environmental health is important to both public health and safety, how involvement in the criminal justice system itself can lead to or worsen health and cognitive problems, and how a public health approach can be successfully applied within the criminal justice system. We contend that just as it is important to identify and take action on the social determinants of health if we want to improve health outcomes, it is equally important to identify and address the social determinants of criminal behavior.

Furthermore, our public health framework sees *social justice* as a foundational cornerstone to public health and safety (see Powers and Faden 2006; Caruso 2016, 2017). In

public health ethics, a failure on the part of public health institutions to ensure the social conditions necessary to achieve a sufficient level of health is considered a grave injustice. An important task of public health ethics, then, is to identify which inequalities in health are the most egregious and thus which should be given the highest priority in public health policy and practice. The public health approach to criminal behavior likewise maintains that a core moral function of the criminal justice system is to identify and remedy social and economic inequalities responsible for crime. Just as public health is negatively affected by poverty, racism, and systematic inequality, so too is public safety. This broader approach to criminal justice therefore places issues of social justice at the forefront. It sees racism, sexism, poverty, and systemic disadvantage as serious threats to public safety and it prioritizes the reduction of such inequalities (see Caruso 2017).

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

Implications

Since legal punishment requires the intentional imposition of a penalty for conduct that is represented as a violation of a law of the state, and since the public health-quarantine model does not involve punishment in this way, we consider it a non-punitive alternative to treatment of criminals. When we quarantine an individual with a communicable disease in order to protect people, we are not intentionally imposing a penalty for illegal conduct. The same is true when we incapacitate the criminally dangerous in order to protect people. The right of self-defense and protection of harm to others justifies the limiting or restricting of liberty, but it does not constitute punishment as standardly understood.

Given that the public health-quarantine model is a non-punitive approach to criminal behavior, a number of important implications follow. First, the model strongly suggests that we favor of a *whole person* approach—one that understands that individuals are embedded in social systems, that criminal behavior is often the result of social determinants, and that prevention is always preferable to incapacitation. Second, after a criminal offense has occurred, courts would need to work with mental health experts, drug treatment professionals, and social service agencies to seek alternatives to incarceration. Lastly, for those who must be incapacitated, they would need to be housed in non-punitive environments designed with the purpose of rehabilitation and reintegration in mind.

Consider, for instance, the physical design of prisons.⁷ With a growing number of exceptions, prisons are harsh, restrictive institutions, designed to enable maximum control over inmates' behavior at any time. As Lutham and Klippan write: "Their scale and appearance instill mistrust and anonymity...The ability to personalise space, have ownership and have personal control over one's situation is intentionally absent. Mostly, these are overtly punitive environments, unlike any other" (2016). These "cold" prison environments have an effect on the

people inside them and they are typically not conducive to rehabilitation and reintegration.

Consider the rates of suicide and self-harm in U.S. prisons. According to the federal Bureau of Justice Statistics, suicides account for more deaths in state and federal prisons than drug and alcohol intoxication deaths, homicide, and accidents combined. Matters are worse in county jails where the suicide rate was 46 per 100,000 in 2013. Incidents of self-harm in England and Wales are also at an all-time high (Ministry of Justice 2016). Furthermore, U.S. and U.K. prisons are also breeding grounds for violence (Bowker 1980; Irwin 1980; Johnson 1987; Ministry of Justice 2016), which is not surprising given that they typically confine large numbers of people in overcrowded quarters and in conditions characterized by material and social deprivation (Bowker 1980; Toch 1992; Toch and Adams 2002; Wolfgang and Ferracuti 1976; Wortley 2005).

In his book *Situational Prison Control* (2005), former prison psychologist Richard Wortley articulates strategies to reduce negative behavior in prison contexts, including through physical design. He suggests (a) setting positive expectations through domestic furnishings that confer trust; (b) reducing anonymity through small prison size; (c) personalizing victims through humane conditions; (d) enabling a positive sense of community through ownership and personalization of the space; and (e) reducing provocation and stress by designing in the capacity for inmates to enact control over environmental conditions and personal space. The current model of U.S. correctional facilities represents the antithesis of each of these strategies. Lutham and Klippan correctly note, “When we create environments that fuel the negative behaviors we naturally associate with criminals, we are caught in a vicious cycle: harsh community and political attitudes toward prisons and prisoners are perpetuated, and overtly punitive prisons continue to be built” (2016).

Some good examples of innovating prison design include Halden Prison in Norway, Leoben in Australia, Enner Mark in Denmark, and the Norwegian prison island of Bastoy. These prisons are purposely designed to reduce crime. Lutham and Klippan explain:

They do this by providing positive opportunities for inmates and building a greater sense of optimism for their future... These spaces are designed to more closely reflect environments in the outside community. The design treats these people not solely as “prisoners” but also as community members—with all the social, vocational and emotional responsibility that this entails. (Lutham and Klippan 2016)

Halden Prison in Norway, for example, features trees throughout its 75-acre site, whereas U.S. prisons are usually devoid of vegetation in order to maximize visibility. In addition, to help inmates develop routines and to reduce the monotony of confinement, designers spread Halden’s living quarters, work areas, and activity centers across the prison grounds. This provides offenders with some degree of autonomy and encourages interpersonal interactions—mirroring the kinds of conditions they will return to upon release. In fact, the Norwegian Correctional Services has officially adopted the *normality principle*—a principle we strongly endorse. The principle is that during the serving of a sentence “life inside [prisons] will resemble life outside as much as possible.”⁸ It further states that, “No one shall serve their sentence under stricter circumstances than necessary for the security of the community. Therefore offenders shall be placed in the lowest possible security regime.” Lastly, it states that prison should be a restriction of liberty but nothing more, that is, “no other rights have been removed by the sentencing court.” According to the normality principle, an offender should have all the same rights as other people living in Norway and life inside should resemble life outside as much as possible. All Norwegian prisoners, for example, have the right to study and they are all allowed to vote. Sentences are

also kept short. On average they are no more than eight months long, and nearly 90% of sentences are for less than a year. Additionally, the longest sentence permitted by law is 21 years, but that can be extended in five-year increments if a prisoner is not rehabilitated and is considered a continued risk to society. Since most prisoners will eventually return to society, Norwegian prisons prepare inmates for reintegration by mimicking the outside world as much as possible.

Adopting the goals of rehabilitation and reintegration will lead not only to reenvisioning the physical design of prisons, but indeed to reconceiving all aspects of the criminal's experience while in custody. The entrenchment of purely punitive approaches to criminal behavior remain a stumbling block in the way of realizing these aims. We believe that in this domain, critical reflection on standard justifications for treatment of criminals is critical to progress.

Conclusion

We have argued that one of the most prominent justifications for legal punishment, retributivism, is inadequate and should be rejected. We began by exploring the retributive justification of legal punishment and explaining why it is inconsistent with free will skepticism. We then outlined our arguments in support of free will skepticism, concluding that it is the position to adopt. We further argue that even if one is not convinced by the arguments for free will skepticism, there are additional reasons for rejecting retributivism, such as the epistemic argument, and concerns about retributivism's dependence on vengeful sentiments. We then introduced a non-retributive and indeed non-punitive alternative for addressing criminal behavior, the public health-quarantine model, which draws on the public health framework and prioritizes prevention and social justice. We concluded with a discussion of some of the implications of this non-punitive approach for criminal justice, focusing on prison design as a case in point.

References

- Balaguer, Mark. (2009) *Free Will As An Open Scientific Problem*, Cambridge, MA: MIT Press.
- Bedau, Hugo Adam, and Erin Kelly. (2015). "Punishment," *Stanford Encyclopedia of Philosophy*, first published on June 13, 2003; substantive revision on July 31, 2015.
<https://plato.stanford.edu/entries/punishment/>
- Berman, Mitchell. (2008). "Punishment and Justification," *Ethics* 18: 258-290.
- Boonin, David. (2008). *The Problem of Punishment*, New York: Cambridge University Press.
- Bowker, L. H. (1980). *Prison Victimization*, Elsevier Science Ltd.
- Caruso, Gregg D. (2012) *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*, Lanham, MD: Lexington Books.
- Caruso, Gregg D. (2016). "Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model," *Southwest Philosophical Review* 32(1): 25-48.
- Caruso, Gregg D. (2017). *Public Health and Safety: The Social Determinants of Health and Criminal Behavior*, UK: ResearchLinks Books.
- Caruso, Gregg D. (2018). "Justice Without Retribution: An Epistemic Argument Against Retributive Criminal Punishment," *Neuroethics*, DOI: 10.1007/s12152-018-9357-8.
- Caruso, Gregg D. (2019). "The Public Health-Quarantine Model," in Dan Nalkin and Derk Pereboom (eds.), *Oxford Handbook of Moral Responsibility*, New York: Oxford University Press.
- Chisholm, Roderick. (1964). "Human Freedom and the Self,: The Lindley Lecture, Department of Philosophy, University of Kansas; reprinted in Gary Watson (ed.), *Free Will*, New York: Oxford University Press.
- Chisholm, Roderick. (1976). *Person and Object*, La Salle IL: Open Court.
- Clarke, Randolph. (2003). *Libertarian Theories of Free Will*, New York: Oxford University Press.
- Corrado, Michael Louis (2017) "Punishment and the Burden of Proof," UNC legal studies research paper. Available at SSRN: <https://ssrn.com/abstract=2997654> or <https://doi.org/10.2139/ssrn.2997654>
- Dennett, Daniel C. (1984). *Elbow Room*, Cambridge, MA: MIT Press.

Dennett, Daniel C., and Gregg D. Caruso. (2018) "Just Deserts: Can we be morally responsible for our actions? Yes, says Daniel Dennett. No, says Gregg Caruso," *Aeon*, October 4, 2018. Available: <https://aeon.co/essays/on-free-will-daniel-dennett-and-gregg-caruso-go-head-to-head>

Double, Richard. 2002. "The Moral Hardness of Libertarianism," *Philo* 5: 226-234.

Ekstrom, Laura W. (2000) *Free Will: A Philosophical Study*, Boulder, CO: Westview.

Feinberg, Joel. (1970) "Justice and Personal Desert," in *Doing and Deserving*, Princeton University Press.

Fischer, John Martin, and Mark Ravizza. (1998) *Responsibility and Control: A Theory of Moral Responsibility*, New York: Cambridge University Press.

Ginet, Carl. (1990) *On Action*, Cambridge: Cambridge University Press.

Griffith, Meghan. (2010). "Why Agent-Caused Actions Are Not Lucky," *American Philosophical Quarterly* 47: 43-56.

Harris, Sam. (2012) *Free Will*, New York: Free Press.

Honderich, Ted. (1988) *A Theory of Determinism*, Oxford: Oxford University Press.

Irwin, John. (1980). *Prisons in Turmoil*, Boston: Little, Brown, and Company.

Johnson, Robert. (1987). *Hard Times: Understanding and Reforming the Prison*, New York: Wadsworth Publishing.

Kane, Robert. (1996) *The Significance of Free Will*, New York: Oxford University Press.

Kant, Immanuel. (1781/1787/1987). *Critique of Pure Reason*, tr. Paul Guyer and Allen Wood, Cambridge, MA: Cambridge University Press.

Kant, Immanuel. (1790). *The Metaphysics of Morals*, Translated by W. Hastie.

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

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

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

Latham, R., and L. Klippan. (2016). "From Expected Reoffender to Trusted Neighbor: Why We

Should Rethink Prisons,” *The Conversation*, August 14: <http://theconversation.com/from-expected-reoffender-to-trusted-neighbour-why-we-should-rethink-our-prisons-60114>

Mele, Alfred. (1995). *Autonomous Agents*, New York: Oxford University Press.

Mele, Alfred. (2006) *Free Will and Luck*, New York: Oxford University Press.

Ministry of Justice. (2016). “Safety in custody statistics bulletin: England and Wales,” <https://www.gov.uk/government/statistics/safety-in-custody-quarterly-update-to-june-2016>

Moore, Michael S. (1987). “The Moral Worth of Retribution,” in Jeffrie G. Murphy (ed.), *Punishment and Rehabilitation*, 3rd Ed., pp.94-130, Wadsworth Publishing Company.

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

Moore, Michael S. (2010). *Placing Blame*, New York: Oxford University Press.

Nelkin, Dana. (2011) *Making Sense of Freedom and Responsibility*, Oxford: Oxford University Press.

O’Connor, Timothy. (2000) *Persons and Causes*, New York: Oxford University Press.

O’Connor, Timothy. (2009) “Agent-Causal Power,” in Toby Handfield (ed.) *Dispositions and Causes*, Oxford: Clarendon Press, pp. 189-214.

Pereboom, Derk. (1995) “Determinism *al dente*,” *Noûs* 29: 21-45.

Pereboom, Derk. (2001). *Living Without Free Will*, Cambridge: Cambridge University Press.

Pereboom, Derk. (2004) “Is Our Conception of Agent Causation Coherent?,” *Philosophical Topics* 32: 275-86.

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

Pereboom, Derk. (2017a) “Responsibility, Agency, and the Disappearing Agent Objection,” *Le Libre-Arbitre, approches contemporaines*, Jean-Baptiste Guillon (ed.), Paris, Collège de France, 2017, pp. 1-18.

Pereboom, Derk. (2017b). “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, Derk, and Gregg D. Caruso (2018). “Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life,” in Gregg D. Caruso and Owen Flanagan (eds),

Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience, pp.1-22, New York: Oxford University Press.

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

Priestley, Joseph. (1788/1963) *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity, in a Correspondence Between Dr. Price and Dr. Priestley*, 1788, Part III: 147-152; reprinted in Joseph Priestley, *Priestley's Writings on Philosophy, Science, and Politics*, John Passmore (ed.) New York: Collie.

Reid, Thomas. (1788/1983) "Essays on the Active Powers of Man," In Sir William Hamilton (ed.) *The works of Thomas Reid*, D.D., ed. Hildesheim: G. Olms Verslagsbuchhandlung, 1983.

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

Shaw, Elizabeth. (2014). *Free Will, Punishment, and Criminal Responsibility*, Dissertation, University of Edinburgh, available at:
<https://www.era.lib.ed.ac.uk/bitstream/handle/1842/9590/Shaw2014.pdf?sequence=2&isAllowed=y>

Smilansky, Saul. (2000) *Free Will and Illusion*, Oxford: Oxford University Press.

Sommers, Tamler. (2007) "The Objective Attitude," *Philosophical Quarterly* 57: 321-41.

Spinoza, Baruch. (1677/1985) *Ethics*, in *The Collected Works of Spinoza*, Edwin Curley, ed. and tr., Volume 1, Princeton NJ: Princeton University Press.

Strawson, Galen. (1986) *Freedom and Belief*, Oxford: Oxford University Press.

Strawson, Peter F. (1962) "Freedom and Resentment," *Proceedings of the British Academy* 48: 187-211.

Taylor, Richard. (1974) *Metaphysics*, Englewood Cliffs: Prentice Hall.

Toch, Hans. (1992). *Living in Prison: The Ecology of Survival*, revised edition, 1985 New York: The Free Press.

Toch, Han, and K. Adams. (2002) *Acting Out: Maladaptive Behavior in Confinement*, Washington, DC: American Psychological Association, APA Books.

van Inwagen, Peter. (1983) *An Essay on Free Will*, Oxford: Oxford University Press.

Vargas, Manuel (2013). *Building Better Beings: A Theory of Moral Responsibility*, New York: Oxford University Press.

Vilhauer, Benjamin. (2013). "Persons, Punishment, and Free Will Skepticism," *Philosophical Studies* 162: 143-63.

Vilhauer, Benjamin. (2015) "Taking Free Will Skepticism Seriously," *Philosophical Quarterly*.

Waller, Bruce. (1990) *Freedom Without Responsibility*, Philadelphia: Temple University Press.

Wolf, Susan. (1990) *Freedom Within Reason*, Oxford: Oxford University Press.

Wolfgang, M.E., and F. Ferracuti. (1976). *The Subculture of Violence*, London: Tavistock.

Wortley, Richard. (2005). *Situational Prison Control*, Cambridge, MA: Cambridge University Press.

Zimmerman, Michael. (2011). *The Immorality of Punishment*, New York: Broadview Press.

¹ E.g., 204 Pa. Code Sect. 303.11 (2005); See also (Tonry 2004).

² E.g., Cal. Penal Code Sect. 1170(a)(1) (West 1985): "The legislature finds and declares that the purpose of imprisonment for crime is punishment."

³ E.g., Model Penal Code Sect. 1.02(2) (Official Draft 1962).

⁴ American Law Institute, Model Penal Code Sect. 1.02(2) adopted May 24, 2017.

⁵ See, for example, the U.S. cases *Spaziano v. Florida*, 468 U.S. 447, 462 (1984); *Gregg v. Georgia*, 428 U.S. 153, 183-84 (1976); Cotton (2000).

⁶ *Spaziano v. Florida*, 468 U.S. at 461.

⁷ The comments below on prison design were first made in Caruso (2017).

⁸ For more details, see the Norwegian Correctional Service's full document:
<http://www.kriminalomsorgen.no/informationin-english.265199.no.html>