1

Free Will, Legal Punishment, and Retributivism

Within the criminal justice system one of the most prominent justifications for legal punishment, both historically and currently, is *retributivism*. The retributive justification of legal punishment maintains that, absent any excusing conditions, wrongdoers are morally responsible for their actions and *deserve* to be punished in proportion to their wrongdoing. Unlike theories of punishment that aim at deterrence, rehabilitation, or incapacitation, retributivism grounds punishment in the *blameworthiness* and *desert* of offenders. It holds that punishing wrongdoers is intrinsically good. For the retributivist, wrongdoers deserve a punitive response proportional to their wrongdoing, even if their punishment serves no further purpose. This means that the retributivist position is not reducible to consequentialist considerations nor in justifying punishment does it appeal to wider goods such as the safety of society or the moral improvement of those being punished.

The dual aims of this book are to argue against retributivism and to develop and defend a viable nonretributive alternative for addressing criminal behavior that is both ethically defensible and practically workable. In the first half of the book, I argue that there are several powerful reasons for rejecting retributivism, not the least of which is that it is unclear that agents possess the kind of free will and moral responsibility needed to justify it. I also consider a number of alternatives to retributivism, including consequentialist deterrence theories, educational theories, and communicative theories, and argue that they have ethical problems of their own. In the second half of the book, I then develop and defend a novel nonretributive approach, which I call the "public health–quarantine model." The model draws on the public health framework and prioritizes prevention and social justice. I argue that it not only offers a stark contrast to retributivism, it also provides a more human, holistic, and effective approach to dealing with criminal behavior, one that is superior to both retributivism and other leading nonretributive alternatives.

1.1 FREE WILL AND THE CRIMINAL LAW

Since the issue of free will features centrally in the discussion to follow, it is important that I begin by defining what I mean by "free will." I maintain that the

variety of free will that has been of central philosophical and practical importance in the historical debate is the sort required for moral responsibility in a particular but pervasive sense. This sense of moral responsibility is typically set apart by the notion of *basic desert* and is defined in terms of the control in action needed for an agent to be truly deserving of blame and praise, punishment and reward. As Derk Pereboom defines it:

> For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary. The desert at issue here is *basic* in the sense that the agent would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations. (2014: 2)

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

There are several distinct advantages to defining free will in this way. First, it provides a neutral definition that virtually all parties can agree to. Unlike some other definitions, it does not beg the question or exclude from the outset various conceptions of free will that are available for disputing parties to adopt. Second, by defining free will in terms of moral responsibility, this definition captures the practical importance of the debate. As Manuel Vargas writes:

> One advantage of making explicit an understanding of free will as linked to responsibility, is that it anchors philosophical concerns in something comparatively concrete and undeniably important to our lives. This is not a sense of free will whose only implication is whether it fits with a given philosopher's particular speculative metaphysics. It is not a sense of free will that is arbitrarily attached to a particular religious framework. Instead, it is a notion of free will that understands its significance in light of the role or function it plays in widespread and recognized forms of life. (2013b: 180)

Third, this definition fits with our everyday understanding of these conceptions. There is, for instance, growing evidence not only that ordinary people view free will and moral responsibility as intimately tied together but also that it is precisely the desire to blame, punish, and uphold moral responsibility that motivates belief in free will (see, e.g., Clark et al. 2014; Shariff et al. 2014; Feldman et al. 2016; Clark et al. 2018; Clark, Winegard, and Baumeister 2019; Everett et al. 2018). People, for

instance, attribute more free will to performers of morally bad actions than morally good actions and morally neutral actions (Feldman et al. 2016; Clark et al. 2018; Everett et al. 2018), and pondering over morally bad actions leads people to increase their belief in free will of all humankind (Clark et al. 2014). They also appear to understand responsibility for actions and consequences first and foremost in the sense defined earlier (Cushman 2008). For instance, empirical findings from Shariff et al. (2014) and others support the hypothesis that free will beliefs, at least among ordinary people, positively predict retributive and backward-looking attitudes (see also Carlsmith and Darley 2008).

Lastly, rejecting this understanding of free will makes it difficult to understand the nature of the substantive disputes that are driving the free will debate. It is important, for instance, to distinguish between *consequentialist-based* and *desert-based* approaches to blame and punishment (see, e.g., Caruso and Morris 2017; Morris 2018; Dennett and Caruso 2021). Consequentialist-based approaches are forward-looking in the sense that agents are considered proper targets of reprobation or punishment for immoral actions on the grounds that such treatment will, say, prevent the agent or other agents from performing that type of action in the future. *Desert-based responsibility*, on the other hand, is considered to be backward-looking and retributivist in the sense that any punitive attitudes or treatments that are deemed appropriate responses to moral or legal offenses are warranted simply by virtue of the action/decision itself, irrespective of whatever good or bad results might follow from the punitive responses. By defining free will as the control in action required for basic desert moral responsibility, we are able to make sense of this distinction and sharpen one of the key questions in the free will debate: Do agents have the control in action (i.e., the free will) needed to justify various desert-based judgments, attitudes, or treatments such as resentment, indignation, moral anger, and retributive punishment? Since those who doubt or deny the existence of free will can adopt consequentialist and forward-looking approaches to blame and punishment, this cannot be what the substantive debate is about (see Caruso and Morris 2017). Hence, whatever else the free will debate is about, it must also be about basic desert moral responsibility.

One of the key claims of this book is that the issue of free will is relevant to the criminal law for at least two main reasons. First, the criminal law is founded on the idea that persons can be held morally responsible for their actions because they have freely chosen them. The US Supreme Court, for instance, has asserted:

> A "universal and persistent" foundation stone in our system of law, and particularly in our approach to punishment, sentencing, and incarceration, is the "belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil."[1]

[1] *United States* v. *Grayson*, 438 U.S. 41 at 52 (1978), quoting *Morissette* v. *United States*, 342 U.S. 246, 250 (1952).

Indeed, the US courts have observed that "[t]he whole presupposition of the criminal law is that most people, most of the time, have free will within broad limits"[2] and that "the law has been guided by a robust common sense which assumes the freedom of the will as a working hypothesis in the solution of its problems."[3] The US Supreme Court, in fact, has gone so far as to suggest that "a deterministic view of human conduct ... is inconsistent with the underlying precepts of our criminal justice system."[4] While this last claim is controversial, since some legal scholars claim that the criminal law only requires compatibilist free will (see Morse 2010, 2013, 2018), one thing is clear: If human beings lack the control in action, that is, the free will, required for basic desert moral responsibility, then our current conception of the criminal law will need to be revised.

As the legal historian Thomas Andrew Green has observed,

> criminal law has been affected by the problem [of free will] at every level: the definition of criminal offenses, the assessment of responsibility, including the practices we have adopted to reach such an assessment; and the way we deal with those found guilty, both in the formal sense of the institutions of punishment or treatment and in the informal sense of social views regarding the guilty (2014: 2–3).

The UK House of Lords has held that "the criminal law generally assumes the existence of free will .... [I]nformed adults of sound mind are treated as autonomous beings able to make their own decisions how they will act."[5] The problem, however, is that the criminal law's appeal to free will is founded more on preference and convenience than fact. As Herbert Packer explains:

> The idea of free will in relation to conduct is not, in the legal system, a statement of fact, but rather a value preference having very little to do with the metaphysics of determinism or free will .... Very simply, the law treats man's conduct as autonomous and willed, not because it is, but because it is desirable to proceed as if it were. (1968: 74–75)

Matthew Jones further writes, "While most areas of law carry free will as a base assumption, criminal law relies on it to an even greater degree because it provides a philosophical basis for individual punishment" (2003: 1035). This brings me to my second point.

Free will is also relevant to the criminal law because of the important role it plays, historically and currently, in one of the most prominent justifications of legal punishment: *retributivism*. The retributive justification of punishment maintains

[2] *Smith* v. *Amontrout*, 865 F.2d 1502, 1506 (8th Cir. 1988).

[3] *Steward Machine Co.* v. *Davis*, 301 U.S. 548. 590 (1937).

[4] *United States* v. *Grayson*, 438 U.S. 41 at 52 (1978). See also *Bethea* v. *United States*, 365 A.2d 64, 83 n.39 (D.C. 1976), which asserts: "[T]he notion that a person's conduct is a simple function of extrinsic forces and circumstances over which he has no control is an unacceptable contradiction of the concept of free will, which is the *sin qua non* of our criminal justice system."

[5] *R* v. *Kennedy*, No. 2 (2007), A.C. 269, at 275.

that 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 when he writes:

> [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. (1997: 153; see also 1987, 1993)

This backward-looking focus on desert is a central feature of most pure retributive accounts of punishment (see, e.g., Kant 1790; von Hirsch 1976, 1981, 2007, 2017; Husak 2000; Kershnar 2000, 2001; Berman 2008, 2011, 2013, 2016; Walen 2014).[6] And it is 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. 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 US criminal justice system, the retributivist justification of legal punishment and the attendant proportionality requirement are widely embraced. In fact, a number of sentencing guidelines in the United States have adopted the retributivist conception of desert as their core principle,[7] and it is increasingly given deference in the "Purposes" section of state criminal codes,[8] where it can be the

[6] In Chapter 5, I will examine so-called mixed accounts of punishment, like R. A. Duff's (2001), that combine both forward- and backward-looking features. For the moment, though, I will focus on unmixed or pure retributive accounts.

[7] E.g., 204 Pa. Code Sect. 303.11 (2005); see also (Tonry 2004).

[8] 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."

guiding principle in the interpretation and application of the code's provisions.[9] Indeed, the American Law Institute recently revised the Model Penal Code so as to set desert as the official dominant principle for sentencing.[10] And courts have identified desert as the guiding principle in a variety of contexts,[11] as with the Supreme Court's enthroning of retributivism as the "primary justification for the death penalty"[12] (Robinson 2008: 145–146). Additional examples can be found in legislation, judicial decisions, sentencing guidelines, and criminal codes in England, Wales, Scotland, Australia, Canada, New Zealand, and Israel (see, e.g., Dingwall 2008; von Hirsch 2017).

For instance, in 2003 the UK Parliament established a Sentencing Guidelines Council consisting of senior judges, legal scholars, and criminologists to provide explicit guidance for sentencing decisions. The following year, the body promulgated an "overarching principle" relying substantially on conceptions of desert (see von Hirsch 2017: fn.9; Sentencing Guideline Council 2004; and Ashworth 2015: Chapter 4). This norm provides that a sentencing court is "required to pass a sentence that is commensurate with the seriousness of the offence," with the offence seriousness to be determined by two main parameters: "the *culpability* of the offender and the *harm* caused or risked being caused by the offence" (2004: s. 1.4; see also von Hirsch 2017: fn.9). Israel also recently adopted a desert-oriented sentencing statue, one that "articulates a retributive guiding philosophy for sentencing and contains a series of provisions related to various aspects of sentencing" (Roberts and Gazal-Ayal 2013: 457). For instance, its Sentencing Act of 2012 stipulates that "[t]he guiding principle in sentencing is proportionality between the seriousness of the offence committed by the offender and the degree of his culpability, and the type and severity of his punishment" (s. 40b). By viewing proportionality to the offender's culpability as the "guiding" principle in criminal sentencing, this provision aligns the statute with sentencing laws in other jurisdictions such as New Zealand and Canada. For example, the Canadian Criminal Code designates proportionality as "the fundamental purpose of sentencing" and states that "[a] sentence must be proportionate to the gravity of the offence and the degree of responsibility of the offender" (2006: s. 718.1). The New Zealand Sentencing Act similarly affirms the importance of proportionality in sentencing (2002: s. 8(a); see also Roberts and Gazal-Ayal 2013: fn.12). And in Australia, one example of judicial attention to the retributive justification of punishment in the context of a sentencing decision can be found in the following comments from Chief Justice Spigelman in a case relating to the torture and murder of a seven-month old child: "[T]he sense of outrage in

[9] E.g., Model Penal Code Sect. 1.02(2) (Official Draft 1962).

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

[11] See, e.g., the US cases *Spaziano* v. *Florida*, 468 U.S. 447, 462 (1984); *Gregg* v. *Georgia*, 428 U.S. 153, 183–184 (1976); Cotton (2000).

[12] *Spaziano* v. *Florida*, 468 U.S. at 461.

the community in such a case is so strong that the element of retribution must play a prominent part in the exercise of sentencing discretion" (*R* v. *Hoerler*, 2014, NSWCCA 184).

Consider, for instance, the recent revisions made by the American Law Institute as part of a fifteen-year project to revise the Model Penal Code, first introduced in 1962. The Model Penal Code is one of the most important developments in American law, and perhaps the most important influence on American criminal law. Conceived as a way to standardize and organize the often fragmentary criminal codes enacted by the states, the Model Penal Code has influenced a large majority of states to change their laws. While the Model Penal Code is not law and has no binding effect, it has been the model for many states' criminal codes and has been extremely influential on state and local lawmakers.

The Purposes provision of the revised code, now set in place by the vote of May 24, 2017, states:

> **§ 1.02(2). Purposes; Principles of Construction.**
> (2) The general purposes of the provisions on sentencing, applicable to all official actors in the sentencing system, are:
> (a) in decisions affecting the sentencing of individual offenders:
> (i) to render sentences in all cases within a range of severity proportionate to the gravity of offenses, the harms done to crimevictims, and the blameworthiness of offenders;
> (ii) when reasonably feasible, to achieve offender rehabilitation, general deterrence, incapacitation of dangerous offenders, restoration of crime victims and communities, and reintegration of offenders into the law-abiding community, provided these goals are pursued within the boundaries of proportionality in subsection (a)(i); and
> (iii) to render sentences no more severe than necessary to achieve the applicable purposes in subsections (a)(i) and (a)(ii).[13]

It is the inclusion of (2)(a)(i) that taints the entire Model Penal Code and sets retributivism as the official dominant principle for sentencing. While (2)(a)(ii) and (2)(a)(iii) are not retributive in nature, they are secondary to the subsection setting the "blameworthiness of offenders" as the primary justification for criminal sentencing. Note, first, that retributivism is to guide sentencing "in all cases," while rehabilitation, deterrence, incapacitation, and restorative justice are only to be pursued "when reasonably feasible." Second, the provision clearly states that (2)(a)(i) trumps (2)(a)(ii) in that the forward-looking, nonretributive approaches cited in (a)(ii) are only to be pursued "within the boundaries of proportionality in

[13] The revised Model Penal Code is available at: https://robinainstitute.umn.edu/sites/robinainstitute.umn.edu/files/mpcs_proposed_final_draft.pdf.

subsection (a)(i)." This amounts to saying that retributive proportional punishment cannot be overridden for forward-looking reasons.

The revised Purposes section therefore rests upon the theory of *limited retributivism*, setting a maximum and minimum for all sentencing based upon deontological and retributive principles, and allowing for forward-looking, nonretributive options only "when reasonably feasible" and "within the boundaries of proportionality." This is unfortunate, I contend, for at least two reasons. First, by setting desert as the official dominant principle for sentencing, the Model Penal Code, along with a number of state sentencing guidelines, has assumed that "most people, most of the time, have free will within broad limits." This assumption is highly questionable and has been the subject of one of the greatest debates in the history of philosophy. It should not be taken for granted. In fact, *free will skepticism* takes seriously the possibility that human beings are never morally responsible in the basic desert sense. Some skeptics defend the moderate claim that in any particular case in which we may be tempted to judge that an agent is morally responsible in the basic desert sense, we lack the epistemic warrant to do so (e.g., Rosen 2004). Others go further and deny that free will and basic desert moral responsibility are even possible (e.g., Strawson 1986, 1994). My own view, and the view I defend in the following chapter, is that our best philosophical and scientific theories about the world indicate 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.

Setting desert as the official dominant principle for sentencing is unfortunate for a second reason as well. As trial judge Michael H. Marcus writes:

> The revision has essentially abandoned any solution other than guidelines to the problem of un-prioritized sentencing purposes, eschewing responsibility for improvement of the public safety performance of sentencing, and settling for whatever guidelines can bring us to moderate mass incarceration and sentencing disparity. Indeed, it is only by proposing guidelines, sentencing commissions, and related appellate review that the revision has any claim to improvement as compared with the [previous] Model Penal Code. Through the various drafts of the revision, it is apparent that the [American Law Institute] believes that programs and alternatives are appropriate for only a small "layer" of crimes …. Because the revision has yielded all to the continued archaic dominance of just deserts – retributivism however named – and because it evades responsibility for public safety, its only promise is that of guidelines. That promise is anemic indeed. (2007: 74–75)

I agree with Judge Marcus that, independent of worries over free will, the move toward retributivism will likely not reduce crime or increase public safety. Focused as it is on just deserts and the blameworthiness of offenders, it makes no effort at addressing the causal determinants of crime or rehabilitating and reintegrating

offenders back into society. On the other hand, the public health approach developed in this book maintains that the criminal justice system should be focused on prevention, rehabilitation, and reintegration. It aims to replace the current *reactive* approach to criminal behavior with a *preventive* approach. I contend that it is indeed unfortunate that the American Law Institute has decided to continue the "archaic dominance" of just deserts and retributivism. When judges and attorneys assume that legal punishment is all (or primarily) about the blameworthiness of offenders and giving them their just deserts, they become blind to the fact that public safety, fairness, and the well-being of society are better served by adopting a more holistic approach focused on prevention, addressing social injustices that give rise to crime, and rehabilitation.

1.2 RETRIBUTIVISM AND PUNISHMENT

Depending on how retributivists view the relationship between desert and punishment, we can identify three different varieties of the view – *weak*, *moderate*, and *strong*.[14] *Weak retributivism* maintains that negative desert, which is what the criminal law is concerned with when it holds wrongdoers accountable,[15] is merely necessary but not sufficient for punishment. That is, weak retributivism maintains that while desert is a necessary condition for punishment, it is not enough on its own to justify punishment – other conditions must also be met. As Alec Walen describes it, weak retributivism is the view that "wrongdoers forfeit their right not to suffer proportional punishment, but that the positive reasons for punishment must appeal to some other goods that punishment achieves, such as deterrence or incapacitation" (2014). Wrongdoing, on this view, is merely a necessary condition for punishment: "The desert of the wrongdoer provides neither a sufficient condition for nor even a positive reason to punish" (Walen 2014; see also Mabbott 1939; Quinton 1954).

Moderate retributivism, on the other hand, maintains that negative desert is necessary and sufficient for punishment but that desert does not mandate punishment or provide an obligation to punish in all circumstances – that is, there may be other goods that outweigh punishing the deserving or giving them their just deserts (Robinson and Cahill 2006). Leo Zaibert, while eschewing the

[14] See, e.g., Alexander, Ferzan, and Morse (2009: 7–10), Walen (2014).

[15] *Negative desert* can be contrasted with *positive desert*, which has to do with an agent deserving praise or reward for good actions. It's important to note that there is another conception of "negative desert" that is widespread in the literature. This latter notion refers to the negative component of the retributivist thesis. As Walen examples, "Retributivism . . . involves both positive and negative desert claims. The positive desert claim holds that wrongdoers morally deserve punishment for their wrongful acts." On the other hand, "[t]his positive desert claim is complemented by a negative one: Those who have done no wrong may not be punished. This prohibits both punishing those not guilty of wrongdoing (who deserve no punishment), and punishing the guilty more than they deserve (i.e., inflicting disproportional punishment)" (2014, s. 3.1). Having two different notions of negative desert can potentially be confusing, but I will try my best to make clear which conception is at play in different contexts.

taxonomy offered here, defends a kind of moderate retributivism when he argues:

> There are many reasons why sometimes refraining from punishing a deserving wrongdoer is more valuable than punishing him – even if one believes that there is [intrinsic] value in inflicting deserved punishment. Perhaps the most conspicuous cases are those in which the refraining is related to resource-allocation and opportunity costs . . . To acknowledge the existence of these cases is not to thereby *deny* the value of deserved punishment: it is simply to recognize that this value, like any value, can be – and often is – lesser than other values. (2018: 20)

Mitchell Berman also defends a form of moderate retributivism, which he calls "modest retributivism" (2016), since he maintains that negative desert grounds a justified reason to punish, but not a duty. For moderate retributivists, negative desert is sufficient to justify punishment but other values and considerations may outweigh inflicting the deserved punishment.

Lastly, *strong retributivism* maintains that desert is necessary and sufficient for punishment but it also grounds a *duty* to punish wrongdoers. Immanuel Kant is perhaps the most famous representative of this latter view, since he famously argued that the death penalty was not only deserved but also obligatory in cases of murder:

> [W]hoever has committed murder, must *die*. There is, in this case, no juridical substitute or surrogate, that can be given or taken for the satisfaction of justice. There is no *Likeness* or proportion between Life, however painful, and Death; and therefore there is no Equality between the crime of Murder and the retaliation of it but what is judicially accomplished by the execution of the Criminal. Even if a civil society resolved to dissolve itself with the consent of all its members – as might be supposed in the case of a people inhabiting an island resolving to separate and scatter themselves throughout the whole world – the last murderer lying in prison ought to be executed before the resolution was carried out. This ought to be done in order that every one may realize the desert of his deeds, and that blood-guiltiness may not remain upon the people; for otherwise they might all be regarded as participators in the murder as a public violation of justice. (1790: Part II: 6)

Of course, not all retributivists support the death penalty – in fact, many contemporary retributivists do not – but here Kant seems to be embodying the strong retributivist view that we are not only justified in giving offenders their just deserts by punishing them, we have a duty to do so. Michael S. Moore also defends a form of strong retributivism and argues, like Kant, that society has a duty to punish culpable offenders:

> We are justified in punishing because and only because offenders deserve it. Moral responsibility ("desert") in such a view is not only necessary for justified punishment, it is also sufficient. Such sufficiency of justification gives society more than merely a right to punish culpable offenders. It does this, making it not unfair to

> punish them, but retributivism justifies more than this. For a retributivist, the moral responsibility of an offender also gives society the *duty* to punish. Retributivism, in other words, is truly a theory of justice such that, if it is true, we have an obligation to set up institutions so that retribution is achieved. (1997: 91)

Strong retributivists therefore defend two distinct claims: (1) that negative desert is sufficient to justify punishing wrongdoers on the grounds that they deserve it and (2) that we have a duty to do so. Moderate retributivists, on the other hand, seek only to defend the first claim.

In what follows, I will limit my discussion to moderate and strong varieties of retributivism and leave weak retributivism aside – although I will briefly return to the weak retributivist thesis in Chapter 4. I will do so because, first, most leading retributivists defend one of these stronger forms of retributivism and it is my desire to address the dominant view, not a subordinate view held by few – see, for example, Moore (1997, 1987, 1993), Kershnar (2000, 2001), Husak (2000), Berman (2008, 2011), von Hirsch (1976, 2007, 2017), Alexander (2013), Alexander, Ferzan, and Morse (2009). Second, weak retributivism is considered by many retributivists to be "too weak to guide the criminal law" and as amounting to nothing more than "desert-free consequentialism side constrained by negative desert" (Alexander, Ferzan, and Morse 2009: 7). In fact, some theorists simply define retributivism in a way that excludes *weak retributivism* from consideration altogether. David Boonin, for example, defines retributivism as the claim that "committing an offense in the past is *sufficient* to justify punishment now, whether or not this will produce any beneficial consequences in the future" (2008: 86; emphasis added). Retributivist Mitchell Berman maintains that the "core retributivist thesis" is that

> [t]he goodness or rightness of satisfying a wrongdoer's negative desert morally justifies [i.e., is sufficient for] the infliction of criminal punishment, without regard for any further good consequences that might be realized as a contingent result of satisfying the wrongdoer's desert. (2016)

And Alec Walen in his *Stanford Encyclopedia of Philosophy* entry on "Retributive Justice" (2014) defines retributivism as committed to the following three principles: (1) that those who commit certain kinds of wrongful acts, paradigmatically serious crimes, morally deserve a proportionate punishment; (2) that it is intrinsically morally good – good without reference to any other goods that might arise – if some legitimate punisher gives them the punishment they deserve; and (3) that it is morally impermissible to intentionally punish the innocent or to inflict disproportionately large punishments on wrongdoers.

Lastly, the weight the criminal law gives desert and the way retributivism is practically implemented in the law (especially in the United States) indicate that the desert of offenders is typically seen as sufficient for punishment. The revised Model Penal Code makes this point rather clear. For these reasons, I will take as my target the claim that the desert of offenders provides sufficient grounds for

punishment and that we are therefore justified in sometimes punishing wrongdoers for no purpose other than to see the guilty get what they deserve. Since this core claim is held in common among all moderate and strong varieties of retributivism, I will henceforth drop the moderate/strong distinction and focus instead on this shared feature.

I will also be limiting my investigation to the question of *legal punishment*. Punishment, we can say, is the intentional imposition of an unpleasant penalty or deprivation for perceived wrongdoing upon a group or individual, typically meted out by an authority. Everyday examples include a parent taking away their teenager's cellphone privileges for a week for bad behavior and a university expelling a student for plagiarism. 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. More precisely, we can say that

> *[l]egal punishment* consists in one person's deliberately harming another on behalf of the state in a way that is intended to constitute a fitting response to some offense and to give expression to the state's disapproval of that offense. (Zimmerman 2011; see also Boonin 2008; Walen 2014)

This is Michael Zimmerman's (2011) definition and it is the one I will adopt herein since I think it best captures the key features of legal punishment. According to this definition, Person A legally punishes Person B if and only if A acts on behalf of the state in such a way that (1) he harms the punishee – this could include imposing an unpleasant penalty or deprivation; (2) this harm is *intended* by the state; (3) this harm is believed by the state to be *fitting* – in particular, fitting to the fact, perhaps in conjunction with some other facts, that the punishee is associated with some legal offense; (4) he thereby expresses the state's disapproval both of the offense and of the offender; and (5) he thereby acts in some legal official capacity (see Zimmerman 2011: 20). Note that the first three conditions set out what we may call the core requirements of punishment: intended harm that is (believed to be) fitting. The latter two conditions are unique to the legal context – that is, legal punishment is distinct from interpersonal punishment or punishment carried out by an angry mob since it is sanctioned by the state, expresses the state's disapproval, and is guided by a system of criminal laws concerned with punishment of individuals who are believed to have committed crimes.

It is important to stress that *intended harm* is a necessary condition of punishment. As Leo Zaibert correctly notes, "[W]hatever else punishment seeks to do, it seeks to make wrongdoers suffer (by somehow diminishing their well-being or by visiting upon them something they do not want)" (2018: 1). He goes on to write:

> To punish . . . is to (try to) inflict suffering (or pain or misery or a bad thing, etc.) on someone as a response to her wrongdoing. Punishment without trying to inflict suffering is like gifting an object without intending to transfer any right over the thing gifted or like feeding someone without intending to give her some nourishment. (Zaibert 2018: 7)

The renowned legal philosopher H. L. A. Hart stresses this point in his definition of punishment, which he maintains "must involve pain or other consequences normally considered unpleasant" (2008: 4). Wittgenstein concurs, finding it perfectly "clear," that just as "reward must be something pleasant," punishment must be "something unpleasant" (1961: 78e; see also Tasioulas 2006; Boonin 2008; Zimmerman 2011; Zaibert 2018). The kind of harm, suffering, or harsh treatment meted out by the state in cases of legal punishment need not involve physical pain, but it must be unpleasant or diminish (at least temporarily) the well-being of offenders.

It must also be the case that the state (or someone acting on its behalf) deliberately intends the harm, since unintended harms do not constitute punishment. If I trip and knock over an elderly person, causing them harm, I do not thereby punish them. Punishment requires *intentional* harm. As Walen writes:

> For an act to count as punishment, it must . . . [f]irst . . . impose some sort of cost or hardship on, or at the very least withdraw a benefit that would otherwise be enjoyed by, the person being punished. Second, the punisher must do so intentionally, not as an accident, and not as a side-effect of pursuing some other end. (2014, s. 2.1)

Zimmerman concurs with this second requirement and writes, "[N]o accidental harming can qualify as punishment; if you punish someone, then the harm that you cause is something that you intended to cause" (2011: 7–8). He goes on to argue:

> This being the case, it might be thought that, although no accidental harm counts as punishment, perhaps the harm that punishment involves needn't be intended either, as long as it is foreseen. But I think that a little reflection shows that that can't be right. We often knowingly cause harm without intending to do so. Indeed, the harm that we knowingly cause might be just the same kind of harm as that caused by punishment but, unless we intend to cause it, we cannot be said to be engaged in punishment. Compare incarceration with quarantine, for example. The extent to which one's liberty is restricted may be the same in either case, but only the former qualifies as punishment, for only in the former case is the harm caused by the restriction to liberty intended. In the latter case, the harm is foreseen but it is not intended – although the restriction itself is of course intended, for otherwise it could not be classified as a case of quarantine at all. (2011: 9–10)

The comparison between punishment and quarantine will play an important role in the discussion to follow, but Zimmerman is absolutely correct that quarantine is not a form of punishment since it does not intentionally seek to cause harm. Punishment, on the other hand, requires that the harms caused be intended.

The question under discussion in this book, then, is whether *retributive legal punishment* is ever justified. That is, does retributivism as a theory of legal punishment withstand scrutiny? In particular, is state-sanctioned intentional imposition of harms (e.g., penalties and sanctions) for violating the state's laws ever basically deserved?

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. 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. Free will skepticism therefore 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. And this would remain true whether or not the criminal law assumes libertarian free will, since free will skepticism also denies that compatibilism preserves the kind of free will and moral responsibility needed to justify retributive legal punishment.

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 2013a) or contractualist (e.g., Ben Vilhauer 2013a, 2013b) considerations. For instance, according to 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 (2013a). 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. Dennett, himself, admits as much in my recent debate with him in *Just Deserts: Debating Free Will* (Dennett and Caruso 2021), where he makes clear that his consequentialist defense of the "system of desert" rejects retributivism. Furthermore, it is unclear that retributivism actually has the consequentialist value some think – this too is an assumption I will challenge in subsequent chapters.

My own reasons for favoring free will skepticism will be spelled out in detail in the following chapter. As we will see, they feature distinct arguments that target three rival views, *event-causal libertarianism*, *agent-causal libertarianism*, and *compatibilism*, and then claim that the skeptical position is the only defensible position that remains standing.[16] I maintain that the sort of free will required for basic desert moral responsibility is incompatible with both causal determination by factors

[16] See Chapter 2 for definitions. But in short, *libertarian* theories of free will reject the thesis of causal determinism and defend an indeterminist conception of free will in order to save what they maintain are necessary conditions for free will – the *ability to do otherwise* in exactly the same set of conditions

beyond the agent's control and with the kind of indeterminacy in action required by the most plausible versions of libertarianism. For this reason, I follow Pereboom in labeling my view *hard incompatibilism*, since it maintains that free will is incompatible with both determinism and indeterminism (Pereboom 2001, 2014; Pereboom and Caruso 2018). I also present a second, independent argument for skepticism that maintains that regardless of the causal structure of the universe, free will and basic desert moral responsibility are incompatible with the pervasiveness of *luck*. This argument is intended not only as an objection to event-causal libertarianism but extends to compatibilism as well. At the heart of the argument is the following dilemma, which Neil Levy (2011) calls the *luck pincer*: Either actions are subject to *present luck* (luck around the time of action), or they are subject to *constitutive luck* (luck in who one is and what character traits and predispositions one has), or both. Either way, luck undermines moral responsibility since it undermines responsibility-level control.

But even if one is not convinced by the arguments against the sort of free will required for basic desert, I argue that there remains a second strong *Epistemic Argument* against causing harm on retributivist grounds that is sufficient for the rejection of retributive legal punishment. 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 (see Vilhauer 2009a, 2012, 2015; Shaw 2014; Corrado 2017; Caruso 2020). 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. This Epistemic Argument against retributive legal punishment is developed and defended in Chapter 3, but I can briefly summarize it here as follows:[17]

and/or the idea that we remain, in some important sense, the *ultimate source/originator* of action. *Compatibilism*, on the other hand, defends a conception of free will that aims to reconcile free will with causal determinism – the thesis that every event or action, including human action, is the inevitable result of antecedent circumstances in accordance with the laws of nature.

[17] The basic idea behind the Epistemic Argument was first introduced by Pereboom in *Living Without Free Will* (2001: 161). Benjamin Vilhauer then developed the argument in detail in a series of papers (2009a, 2012, 2015). I presented my own version of it in Caruso (2020). Others who have also defended versions of the Epistemic Argument include Double (2002), Shaw (2014), Corrado (2017), and Jeppsson (2020).

(1) Legal punishment intentionally inflicts harms on individuals and the justification for such harms must meet a high epistemic standard. If it is significantly probable that one's justification for harming another is unsound, then, prima facie, that behavior is seriously wrong.
(2) The retributivist justification for legal punishment assumes that agents are morally responsible in the basic desert sense and hence justly deserve 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 faces powerful and unresolved objections and as a result falls 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.

In developing the Epistemic Argument, I examine the leading extant accounts of basic desert moral responsibility and argue that they all fail to satisfy the high epistemic standard needed to justify retributive legal punishment – a standard, I contend, that is both reasonable and defensible. I therefore conclude that the Epistemic Argument provides sufficient reason for rejecting retributivism.

In addition to free will skepticism and the Epistemic Argument, I also provide several additional arguments against retributivism in Chapter 4. First, I argue that, even if the requisite capacity for control is in place and basic desert could be secured, there are good pragmatic reasons for rejecting retributivism. This is because retributive justice has "limited effectiveness in promoting important social goals such as rehabilitation and reforming offenders" (Shaw 2019a). Second, I argue that it is philosophically problematic to impart to the state the function of intentionally harming wrongdoers in accordance with desert since it is not at all clear that the state is capable of properly tracking the desert and blameworthiness of individuals in any reliable way. This is because criminal law is not properly designed to account for all the various factors that affect blameworthiness, and as a result the moral criteria of blameworthiness is often misaligned with the legal criteria of guilt (Kelly 2018). While a retributivist could argue that this second objection is more indicative of our broken criminal justice system, not retributivism per se, I argue that there are in principle reasons to doubt that the criminal justice system could ever properly align the legal criteria of guilt with the moral criteria of blameworthiness.

This is because, third, for the state to be able to justly distribute legal punishment in accordance with desert, it needs to be in the proper epistemic position to

know what an agent basically deserves, but since the state is (almost) never in the proper epistemic position to know what an agent basically deserves, it follows that the state is not able to justly distribute legal punishment in accordance with desert. In Chapter 4, I label this the *Poor Epistemic Position Argument* (or PEPA) so as to distinguish it from the Epistemic Argument outlined earlier. I argue that it is practically impossible for the state to punish in accordance with desert, *without this leading to injustice*, since, due to the state's epistemic limitations, some people will inevitably end up getting more punishment than they deserve (and some less) judged on the retributivists own grounds. Lastly, I argue that even if retributivism could overcome all these various difficulties, there remains the additional problem that how the state goes about judging the gravity of wrong done, on the one hand, and what counts as proportional punishment for that wrong, on the other, is wide open to subjective and cultural biases and prejudices, and as a result, the principle of proportionality in *actual practice* does not provide the kind of protections against abuse it promises.

In the end, I conclude that there are at least six powerful reasons for rejecting retributivism: (1) the truth of free will skepticism (or what we can call the *Skeptical Argument*); (2) the *Epistemic Argument*; (3) the *Limited Effectiveness Argument* – that is, practical concerns over the limited effectiveness of retributive justice in promoting important social goals; (4) the *Misalignment Argument*, which argues that the criminal law is not properly designed to account for all the various factors that affect blameworthiness, and as a result the moral criteria of blameworthiness is often misaligned with the legal criteria of guilt; (5) the *Poor Epistemic Position Argument* (or PEPA); and (6) the *Indeterminacy in Judgment Argument*, which argues that judgments of gravity and proportionality are indeterminate, ambiguous, and too easily influenced by cultural biases and prejudices to guarantee respect for persons and prevent cruel and inhumane punishment.

Consider, for instance, the fact that the US criminal justice system widely embraces the retributivist justification of legal punishment and the proportionality requirement, yet this has not prevented the mass incarceration of its citizens nor has it been sufficient to protect against disproportionate punishment. While I am not suggesting that these problems are due solely to retributivism – since there is good reason to think that "three strikes" laws, mandatory minimums, and other policies responsible for mass incarceration were also motivated by the desire for deterrence[18] – my point is simply that commitment to limited retributivism has not *prevented* these abuses. As witnessed by the codes, provisions, and rulings cited earlier, it is clear that the retributivist justification of legal punishment and the attendant proportionality requirement are widely embraced in the United States (see fns.5–10). Yet, with only 4.5 percent of the world's population, the United States imprisons 25 percent of the world's prisoners – far more than any other nation in the world.

[18] This is partly why I also reject consequentialist deterrence theories of punishment (see Chapter 5).

The US imprisons roughly 700 prisoners for every 100,000 people,[19] whereas Scandinavian countries such as Sweden, Finland, and Norway hover around 60 per 100,000.[20] And not only does the United States imprison at a much higher rate, it also imprisons in notoriously harsh conditions.

American prisons are often cruel places, using a number of harsh forms of punishment including extended solitary confinement. The watchdog organization Solitary Watch estimates that 80,000 to 100,000 people in the United States are currently in some form of solitary confinement.[21] These prisoners are isolated in windowless, soundproof cubicles for 23 to 24 hours each day, sometimes for decades. Such excessively punitive punishment not only causes severe suffering and serious psychological problems, it does nothing to rehabilitate prisoners nor does it reduce the rate of recidivism. In fact, the United States has one of the highest rates of recidivism in the world, with a recent report by the Bureau of Justice Statistics (2018) finding an estimated 68 percent of released prisoners being rearrested within three years of release, 79 percent within six years, and 83 percent within nine years.[22] Norway, by contrast, has the world's lowest recidivism rate at 20 percent (Kristoffersen 2010; Encartele Inc. 2018). One of the big differences is that the Norwegian systems aim at rehabilitation and reintegration and prepare inmates for life after incarceration by providing them with educational opportunities and work training. It also encourages guards to cultivate friendships with the inmates, for inmates to interact with each other, and for inmates to exercise maximal autonomy. Treating inmates humanely helps them develop the interpersonal skills needed to successfully reintegrate back into society. US prisons, on the other hand, tend to isolate inmates and control every aspect of their lives – for example, when they eat and sleep, what they do, where they go, etc. This can instill a kind of learned helplessness that makes it extremely difficult for individuals to readjust to life on the outside.

Sentences in the United States also tend to be more punitive than other countries. Even controlling for crime rates and population size, the United States hands down longer sentences, spends more money on prisons, and executes more of its citizens than every other advanced industrial democracy (Farrell and Clark 2004; Blumstein, Tonry, and Van Ness 2005; Enns 2006: 3; Cowen 2010; Amnesty International 2012). A combination of three strikes laws and other legislation has also led to a disproportionate number of US prisoners serving life without the chance of parole.

[19] According to the Prison Policy Initiative, in 2018 the number was 698 per 100,000 (see: www.prisonpolicy.org/global/2018.html).

[20] According to the latest numbers, Norway imprisons 63 per 100,000, Sweden 59, Finland 51, and Denmark 63. See International Centre for Prison Studies, "World Prison Brief": www.prisonstudies.org/sites/default/files/resources/downloads/wppl_12.pdf.

[21] See also the Vera Institute report on "Solitary Confinement: Common Misconceptions and Emerging Safe Alternative" (2015). Available at: www.vera.org/downloads/publications/solitary-confinement-misconceptions-safe-alternatives-report_1.pdf.

[22] The report is available here: www.bjs.gov/content/pub/pdf/18upr9yfup0514.pdf.

In fact, the proportion of individuals serving life without parole in the United States is approximately 180 *times* greater than England (Enns 2016: 3). A recent report by the Sentencing Project (2017) found that there were 161,957 people serving life sentences, or one of every nine people in prison. An additional 44,311 individuals are serving "virtual life" sentence, yielding a total population of life and virtual life sentences at 206,268 – or one of every seven people in prison. The pool of people serving life sentences has more than quadrupled since 1984. Furthermore, nearly half (48.3 percent) of life and virtual life-sentenced individuals are African American, equal to one in five black prisoners overall. More than 17,000 individuals with life without parole or virtual life sentences have been convicted of nonviolent crimes. The United States incarcerates people for life at a rate of 50 per 100,000, roughly equivalent to the entire incarceration rates of the Scandinavian nations of Denmark, Finland, and Sweden (The Sentencing Project 2017).

Of course, there are many reasonable retributivists who acknowledge that we imprison far too many people, in far too harsh conditions. The problem, however, is that retributivism remains committed to the core belief that criminals *deserve* to be punished and suffer for the harms they have caused. This retributive impulse in *actual practice* (rather than in pure theory) often leads to practices and policies that try to make life in prison as unpleasant as possible. It also fuels the desire to continue to punish individuals even after they have been released. Consider voter disenfranchisement. In the 2016 US Presidential election, an estimated 6.1 million Americans were barred from participating due to felony convictions, with an estimated 3.1 million of those disenfranchised due to state laws that restrict voting rights even after the completion of sentences. Such voter disenfranchisement policies can only be justified on the assumption (correct or not) that it is *deserved*, since denying the vote to ex-offenders accomplishes little of forward-looking value. As Supreme Court Justice Thurgood Marshall so eloquently put it:

> It is doubtful . . . whether the state can demonstrate either a compelling or rational policy interest in denying former felons the right to vote. [Ex-offenders] have fully paid their debt to society. They are as much affected by the actions of government as any other citizen, and have as much of a right to participate in government decision-making. Furthermore, the denial of a right to vote to such persons is a hindrance to the efforts of society to rehabilitate former felons and covert them into law-abiding and productive citizens.[23]

Since retributivists care little about rehabilitation and other forward-looking goods, the only rationale they need appeal to is giving wrongdoers their just deserts.[24]

[23] *Richardson* v. *Ramirez*, 418 U.S. at 78; Marshall J. dissenting

[24] Other potential policy interests defenders of voter disenfranchisement can point to include: (1) protecting against voter fraud or other election offenses, (2) prevention of harmful changes to the law, and (3) protection of the "purity" of the ballot box. But there are severe problems with each of these putative interests (see Human Rights Watch, Sentencing Project: www.hrw.org/legacy/reports98/vote/usvot980-03.htm).

Unfortunately, this desire for retribution, in actual practice, often overrides any theoretical commitments to proportionality and respecting human dignity.

1.3 THE PUBLIC HEALTH–QUARANTINE MODEL

If we reject retributivism, either because we come to doubt or deny the existence of free will or for other reasons, what are we left with? Many worry that without the justification of retributivism and the putative protection afforded by the principle of proportionality, we would be unable to successfully deal with criminal behavior. But traditionally, in addition to pure retributivism there have been a number of other common justifications of legal punishment, including consequentialist deterrence theories, moral education theories, and a variety of expressive, communicative, and mixed theories of punishment. While I will argue in Chapter 5 that these other approaches face significant moral concerns of their own – or, in the case of mixed theories, retain certain retributive components that are unjustified – the second half of the book will develop and defend what I contend is an ethically defensible and practically workable alternative to retributive legal punishment, one that is consistent with free will skepticism and preferable to these other nonretributive alternatives. I call it the *public health–quarantine model*. As we will see, the model not only provides a justification for the incapacitation of dangerous criminals consistent with free will skepticism, it also provides a broader and more comprehensive approach to criminal behavior generally since it draws on the public health framework and prioritizes prevention and social justice.

The core idea of the model is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. Yet the model does not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. In fact, the model is completely nonpunitive since it does not satisfy the definition of punishment introduced earlier. It also requires special attention to the well-being and dignity of criminals that would change much of current policy. Perhaps most importantly, the model also develops a public health approach that prioritizes prevention and social justice and aims at identifying and taking action on the social determinants of health and criminal behavior (see Chapters 6–8).

The model begins with Derk Pereboom's quarantine analogy, which 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; Caruso and Pereboom 2020). 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 nonresponsible carrier of a serious communicable disease (see Pereboom 2001, 2014).

The first thing to note about the theory is that although one might justify quarantine (in the case of disease) and incapacitation (in the case of dangerous criminals) on purely utilitarian or consequentialist grounds, Pereboom and I resist this strategy (see Pereboom and Caruso 2018; Caruso and Pereboom 2020). Instead, we maintain 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 nonretributive options (see, e.g., Pereboom and Caruso 2018). One advantage it has, say, over consequentialist deterrence theories is that it has more restrictions placed on it with regard to using people merely as a means. For instance, as it is illegitimate to treat carriers of a disease more harmfully than is necessary to neutralize the danger they pose, treating those with violent criminal tendencies more harshly than is required to protect society will be illegitimate as well. In fact, the model requires that we adopt the *principle of least infringement*, which holds that the least restrictive measures should be taken to protect public health and safety. This ensures that criminal sanctions will be proportionate to the danger posed by an individual, and any sanctions that exceed this upper bound will be unjustified.

Second, the quarantine model places several constraints on the treatment of criminals (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 this model the use of incapacitation should be limited to only those cases where offenders are a serious threat to public safety and no less restrictive measures were available. In fact, for certain minor crimes perhaps only some degree of monitoring could be defended. Secondly, the incapacitation account that results from this analogy demands a degree of concern for the rehabilitation and well-being of the criminal that would alter much of current practice. Just as fairness recommends that we seek to cure the diseased we quarantine, so fairness would counsel that we attempt to rehabilitate the criminals we detain (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 we will see, I situate 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 Chapter 7).

Since the *social determinants of health* and the *social determinants of criminal behavior* are broadly similar, or so I will argue, the best way to protect public health and safety is to adopt a public health approach for identifying and taking action on these shared social determinants. Such an approach requires investigating how social inequities and systemic injustices affect health outcomes and criminal behavior, how poverty affects health and incarceration rates, how offenders often have preexisting medical conditions including mental health issues, how homelessness and education affect 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 (see Chapter 7). I 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, the public health framework sees *social justice* as a foundational cornerstone to public health and safety (see Powers and Faden 2006). In public health ethics, a failure on the part of public health institutions to ensure the social conditions necessary to achieve a sufficient level of health is considered a grave injustice. An important task of public health ethics, then, is to identify which inequalities in health are the most egregious and thus which should be given the highest priority in public health policy and practice. The public health approach to

criminal behavior likewise maintains that a core moral function of the criminal justice system is to identify and remedy social and economic inequalities responsible for crime. Just as public health is negatively affected by poverty, racism, and systematic inequality, so too is public safety. This broader approach to criminal justice therefore places issues of social justice at the forefront. It sees racism, sexism, poverty, and systemic disadvantage as serious threats to public safety and it prioritizes the reduction of such inequalities.

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, I maintain, is sufficient for dealing with dangerous criminals, leads to a more humane and effective social policy, and is actually preferable to the harsh and often excessive forms of punishment that typically come with retributivism.

1.4 IMPLICATIONS

Adopting the public health–quarantine model will, of course, have significant implications for our institutions of criminal justice and law. Since legal punishment seeks to make wrongdoers suffer and 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, it offers a nonpunitive alternative to treatment of criminals. When we quarantine an individual with a communicable disease in order to protect people, we are not intentionally seeking to harm or impose a penalty on them. The same is true when we incapacitate the criminally dangerous in order to protect society. The right of self-defense and prevention of harm to others justifies the limiting or restricting of liberty, but it does not constitute punishment as standardly understood. This is important for several reasons. First, the model demands that we view individuals holistically and that we adopt a preventive 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 nonpunitive environments designed with the purpose of

rehabilitation and reintegration in mind. Since most prisons in the United States, United Kingdom, and Australia are inhospitable and unpleasant places designed for punitive purposes, we would need to redesign them so that the physical environments and spaces we incapacitate people in better serve the goal of rehabilitation and reintegration (see Chapters 7 and 8 for details).

With regard to sentencing, voter disenfranchisement, and three strikes laws, the public health–quarantine model would require radical changes to our current, often excessively punitive, system. For one, felony voter disenfranchisement policies would be ended, since they serve no forward-looking benefit and are ultimately "a hindrance to the efforts of society to rehabilitate former felons and covert them into law-abiding and productive citizens." The sentence of life without parole would also be removed as an option, since it precludes from the outset the possibility of rehabilitation, violates the principle of least infringement, discourages the state from working to improve the well-being of offenders since they are seen as "lost causes," and prevents the reassessment of individual cases as circumstances change. While this may sound like a radical proposal, it is not a novel one. Portugal was the first country to abolish life imprisonment in 1884 and since then several other countries have followed suit, including Norway, Spain, Canada, Bosnia, Brazil, and Croatia. The maximum sentence one can get in Norway, for instance, is twenty-one years, and this is true even for murder. The Norwegian penal code does allow, however, for detention to be extended by five years at a time if the offender is still considered dangerous after serving their original sentence. While this may effectively allow for a life term if an offender cannot be rehabilitated and is deemed a serious continued threat to society, it has the benefit of requiring continued reevaluation and assessment. And unlike the United States, where one out of every nine people in prison is serving a life sentence, very few people in Norway ever serve more than a year. The average sentence in Norway is around eight months, more than 60 percent of prison sentences are up to three months, and almost 90 percent are less than a year.

The public health–quarantine model would similarly require incapacitation be kept to an absolute minimum and life without parole be eliminated as an option. A recent study by the Brennan Center for Justice found that 39 percent of the US prison population (roughly 576,000 people) is behind bars with little public safety rationale (Austin et al. 2016). They found that many offenders should not have been sent to prison in the first place. According to the report, up to 25 percent of prisoners (364,000 people), almost all nonviolent, lower-level offenders, would be better served by alternatives to incarceration such as treatment, community service, and probation. Another 14 percent (212,000 people) have already served long sentences and can be safely set free with little or no risk to public safety. Releasing these individuals would significantly and safely cut our prison population by more than a half million *and* save $20 billion annually – enough to employ 327,000 school teachers, 360,000 probation officers, or 270,000 new police

officers. If these numbers are correct, adopting the public health–quarantine model would significantly aid in reducing mass incarceration since it would end life without parole and seek alternatives to incapacitation for offenders who pose little to no risk to public safety.

Three strikes laws would likewise need to be reversed since they prevent individual cases from being judged on their own terms, which would be needed if we are to accurately assess the threat an individual poses to society moving forward. Not all felonies are equal. And the fact that one person has committed three strikes does not mean they represent the same threat to society as another. Furthermore, three strikes laws often run afoul of principle of least infringement. Consider the following example:

> On November 4, 1995, Leandro Andrade walked into a Southern California Kmart. Andrade – who had several past criminal convictions – was about to commit a crime that would lead to a prison sentence of twenty-five years to life. Two weeks later, still a free man, Andrade struck again. This time, the target was a Kmart just three miles to the west of his previous crime. His plan was identical and would result in another sentence of twenty-five years to life. In two weeks, Andrade had attempted to steal nine VHS tapes: *The Fox and the Hound*, *The Pebble and the Penguin*, *Snow White*, *Batman Forever*, *Free Willy 2*. *Little Women*, *The Santa Clause*, and *Cinderella*. The total cost of the movies was $153.54. The actual cost to Andrade was fifty years to life. (Enns 2016: 1)

Under California's three strikes laws at the time, two counts of petty theft with a prior conviction carried consecutive sentences of twenty-five years to life. And while you might think this an extreme example, it is not:

> In 2010, 32,392 individuals were imprisoned in California [alone] with their second strike. Another 8,764 were incarcerated with their third strike. Of the second strikers, 833 were for petty theft. And like Andrade . . ., an additional 341 individuals faced a potential life sentence for stealing items valued at $950 or less. (Enns 2016: 2)

On the public health–quarantine model, sentences like this would be in violation of the principle of least infringement since Leandro Andrade's acts of petty theft do not constitute a serious threat to society, at least not on their own, and should therefore be dealt with by other, less restrictive means.

These are only a few examples of the kinds of implications the public health–quarantine model would have, but hopefully they provide a sense of the wide-ranging nature of the reforms that would be required. Other implications will be discussed in subsequent chapters.

1.5 COMMON MISCONCEPTIONS

Since one of the main goals of this book is to develop and defend the public health–quarantine model, I would like to address a few potential misconceptions at the outset. Critics sometimes fear that my proposal will place offenders outside the

moral community or flatten out important distinctions, treating the different capacities of agents as irrelevant and/or treating all offenders as if they were incompetent or mentally ill. For instance, retributivist C. S. Lewis (1953) famously argued that if retributivism were abandoned in favor of harm prevention, responsible offenders would be objectified and would no longer be protected by considerations of justice. According to Lewis:

> The Humanitarian theory removed from Punishment the concept of Desert. But the concept of Desert is the only connecting link between punishment and justice . . . There is no sense in talking about a "just deterrent" or a "just cure." We demand of a deterrent not whether it is just but whether it will deter. We demand of a cure not whether it is just but whether it succeeds. Thus when we cease to consider what the criminal deserves and consider only what will cure him or deter others, we have tacitly removed him from the sphere of justice altogether; instead of a person, a subject of rights, we now have a mere object, a patient, a "case." (1953: 225)

In recent years, similar concerns have been echoed by Morse (2018), Smilansky (2011),and Dennett (2011, Dennett and Caruso 2020). Sometimes the concern is put in terms of P. F. Strawson's (1962) famous distinction between the *reactive* and *objective* attitudes. Strawson maintained that our justification for claims of blameworthiness and praiseworthiness is grounded in the system of human reactive attitudes, such as moral resentment, indignation, guilt, and gratitude. Strawson contends that because our moral responsibility practices are grounded in this way, if we were to abandon the notion of moral responsibility, we would be forced to adopt the cold and calculating objective attitude toward others, a stance that relinquishes the reactive attitudes. According to Strawson and his followers, the denial of all moral responsibility is unacceptable, self-defeating, and/or impossible, since to permanently excuse everyone would entail that "nobody knows what he's doing or that everybody's behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense" (1962: 74).

While I take these concerns seriously and acknowledge that earlier versions of what Lewis (1953) calls the "humanitarian" approach may have been guilty of them, I also maintain that they are misguided when applied to the public health–quarantine model I defend. First, it is important to recognize that neither Pereboom nor I set out our position in a strict consequentialist theoretical context. Rather, we justify incapacitation on the ground of the right to self-defense and defense of others. That right does not extend to people who are nonthreats. The aim of protection is justified by a right with clear bounds and not by a consequentialist theory on which the bounds are unclear. It would therefore be wrong, according to our model, to incapacitate someone who is innocent since they are not a serious threat to society (Pereboom and Caruso 2018; see also Chapter 8). The principle of least infringement would also prohibit legal sanctions that exceed what is needed to protect

public health and safety. As a result, it would oppose using individuals simply as a means to deter others by ratcheting up various punitive responses to crime, as was the case with three strikes laws and mandatory minimums. The public health–quarantine model therefore has distinct advantages over consequentialist theories of punishment. Furthermore, Victor Tadros (2011) has persuasively argued that "the objection against using someone merely as a means is best characterized as an objection against '*manipulative use*' – where someone is used in order to promote some further, independent goal. In contrast, harming someone to *eliminate* a threat they pose is much easier to justify, based on the right to self-defense" (Shaw 2019a: 101). Assuming, for the moment, that Tadros's distinction between eliminating harm and manipulative use is valid, concerns over treating individuals merely as means to an end, or as objects to be used, would appear to apply more to consequentialist theories of punishment than to the public health–quarantine model.

Second, neither free will skepticism nor the public health–quarantine model implies that the difference between agents who are reasons-responsive and those who are not is irrelevant to how we should treat offenders. On the contrary, free will skeptics typically claim that this difference is crucial for determining the right response to crime (Pereboom and Caruso 2018; Focquart et al. 2020). On the public health–quarantine model, the question of whether an offender is reasons-responsive is relevant for at least two reasons. First, it is important in assessing what kind of threat an individual poses moving forward and whether incapacitation is required. An offender suffering from a serious mental illness, for instance, differs in significant ways from one who is fully reasons-responsive. These differences will be relevant to determining what minimum restrictions are required for adequate protection. Second, if the capacities of reasons-responsiveness are in place, forms of treatment that take rationality and self-governance into account are appropriate. On the other hand, those who suffer from impairments of rationality and self-governance would need to be treated differently and in ways that aim to restore these capacities when possible. Understanding the variety of causes that lead to impairment of these capacities would also be crucial to determining effective policy for recidivism reduction and rehabilitation (Focquart et al. 2019). Hence, free will skepticism and the public health–quarantine model acknowledge the importance of reasons-responsiveness, self-governance, and differences in degrees of autonomy. But rather than see these as relevant for assigning blameworthiness and basic desert moral responsibility, they instead view them as important (in fact, essential) to determining the appropriate course of action moving forward.

Third, while it is true that in justifying incapacitation the public health–quarantine model appeals to an analogy with quarantine, it is also important to recognize that what is analogous here is the *justification* of incapacitation and quarantine – that is, both appeal to the right of self-defense and prevention of harm to others. The model does not require us to view wrongdoers as "ill" or "diseased." While some people who commit criminal acts do so because of mental illness or incompetency,

many others do not. These distinctions matter exactly for the reasons just stated. Furthermore, I strongly want to resist what Bruce Waller (2006, 2011) calls "excuse-extensionism" – the idea that the denial of moral responsibility only makes sense on the basis of characteristics that make one incompetent (and thus excused) as a moral being (2011: 219). As Waller explains:

> [I]f we start from the assumption of the moral responsibility system (assumptions that are so common and deep that they are difficult to escape), then the denial of moral responsibility is absurd and self-defeating. But the universal denial of moral responsibility does *not* start from the assumption that under normal circumstances we are morally responsible, and it does *not* proceed from that starting point to enlarge and extend the range of excuses to cover everyone (so that *everyone* is profoundly flawed). That is indeed a path to absurdity. Rather, those who reject moral responsibility reject the basic system which starts from the assumption that all minimally competent persons (all who reach the plateau level) are morally responsible. For those who deny moral responsibility, it is never fair to treat anyone as morally responsible, no matter how reasonable, competent, self-efficacious, strong-willed, and clear-sighted that person may be. (2015: 103)

Since skeptics who globally challenge the moral responsibility system do not accept the rules of that system,[25] it would be wrong to interpret them as claiming that "nobody knows what he's doing or that everybody's behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense" (Strawson 1962: 74). Instead, global skeptics maintain that our ordinary moral responsibility practices are unjustified *even when agents are competent and reasons-responsive*.

Fourth, rejecting basic desert moral responsibility does not require one to adopt what Strawson calls the objective attitude toward wrongdoers. As Pereboom and I have elsewhere argued:

> Strawson may be right to contend that adopting the objective attitude would seriously hinder our personal relationships (for a contrary perspective, see Sommers 2007). However, a case can be made that it would be wrong to claim that this stance would be appropriate if determinism did pose a genuine threat to the reactive attitudes (Pereboom 1995, 2001, 2014). While, for instance, kinds of moral anger such as resentment and indignation might be undercut if free will skepticism were true, these attitudes may be suboptimal relative to alternative attitudes available to us, such as moral concern, disappointment, sorrow, and moral resolve. The proposal is that the attitudes that we would want to retain either are not undermined by a skeptical conviction because they do not have presuppositions that conflict with this view, or else they have alternatives that are not under threat. And what remains does not amount to Strawson's objectivity of attitude and is sufficient to sustain the personal relationships we value. (Pereboom and Caruso 2018: 201)

[25] See, e.g., Waller (2011, 2015), Pereboom (2001, 2014), Levy (2011), G. Strawson (1986, 1994), and Caruso (2012, 2013, 2018, 2019).

According to *optimistic skeptics*, like myself, life without belief in free will and basic desert moral responsibility would not be as destructive as Strawson and Lewis contend.[26] Instead, prospects of finding meaning in life and sustaining good interpersonal relationships would not be threatened (see Pereboom 2001, 2014; Waller 1989, 2011, 2014; Caruso 2013, 2018, 2019). And although retributivism and sever punishment, such as the death penalty, would be ruled out, incapacitation and rehabilitation programs would still be justified.[27]

Lastly, critics are simply mistaken when they assume that adopting the skeptical perspective means abandoning all discourse about justice, rights, and respect for persons.

In the second half of the book, I argue that free will skepticism and the public health–quarantine model are perfectly consistent with respect for persons as well as various nondesertist theories of justice. In particular, I defend a *capabilities approach* to social justice and argue that not only is it consistent with free will skepticism, it can also serve as the moral foundation for my public health framework (Chapter 6). Appealing to work by Stephen Darwall (1992) and Benjamin Vilhauer (2013a, 2013b), I also argue that respect for persons and other safeguards against manipulative use can be made consistent with free will skepticism and can be incorporated into the public health–quarantine model (see Chapters 6 and 8; see also Shaw 2019a). For instance, the public health–quarantine model incorporates several important safeguards, including the *conflict resolution principle*, the *principle of least infringement*, the *principle of normality*, and the *prohibition of manipulative use*. There is no reason to think, then, that adopting the public health–quarantine model would require treating individuals as objects. Quite the opposite. I contend that, contra retributivists, we should resist narrowly conceiving of respect for persons in terms of giving wrongdoers their just deserts. Instead, respecting human dignity demands, at a minimum, that the capabilities and well-being of wrongdoers be taken into consideration, that we avoid punishments that dehumanize and disenfranchise individuals, and that we do everything we can to rehabilitate and reintegrate offenders back into the community. The public health–quarantine model does a better job at respecting human dignity in this sense than does retributivism (see Chapter 8).

[26] By "optimistic skeptics" I mean those who are optimistic about the implications of adopting the skeptical perspective (e.g., Pereboom 2001, 2014; Waller 1989, 1990, 2007, 2011, 2015; Nadelhoffer 2011; Levy 2011, 2015, 2016; Caruso 2013, 2018, 2019; Shaw 2019a, 2019b; Vilhauer 2009b, 2012, 2013a, 2013b; Morris 2018).

[27] Other free will skeptics who defend either incapacitation accounts or forward-looking consequentialist accounts include Pereboom (2001, 2013, 2014), Jones (2003), Vilhauer (2009b, 2013a, 2013b), Levy (2012), Corrado (2013), Chiesa (2011), Caruso (2016, 2017, 2020), Pereboom and Caruso (2018), Shaw (2014, 2019a, 2019b), Focquaert, Glenn, and Raine (2013, 2018), and Focquaert (2019). See also Shaw, Pereboom, and Caruso (2019).

1.6 BELIEF IN FREE WILL REMAINS AN OBSTACLE

I would like to close this chapter with some final thoughts on why the issue of free will should not be overlooked in discussions of criminal justice. It has been suggested to me, by friend and foe alike, that if my ultimate goal is criminal justice reform and the rejection of retributive practices and policies, I should work directly toward that goal and leave the contentious issue of free will aside. While I acknowledge that this may be a better rhetorically strategy, since getting people to question the existence of free will is no easy task, I resist it for two main reasons. First, I am a philosopher and am committed to the truth wherever it leads me. While this might sound high-minded, I have worked on the problem of free will for many years and am convinced that free will skepticism is the only reasonable position to adopt. This is not to say, of course, that the skeptical perspective is not consistent with other conceptions of responsibility – for example, causal responsibility, "take charge" responsibility, attributability, answerability, etc. (see Caruso 2018). Nor is it to deny that there remain good reasons for incapacitating dangerous criminals and engaging in forms of moral protest in the face of bad behavior (see Pereboom 2014, 2020). Rather, it is to insist that to hold people *truly deserving* of blame and praise, punishment, and reward would be to hold them responsible for the results of the morally arbitrary or what is ultimately beyond their control, which is fundamentally unfair and unjust. Given that I am convinced of the skeptical perspective, it would be utterly disingenuous of me to set the issue of free will aside or simply ignore it.

Second, and more to the point, to achieve meaningful and lasting criminal justice reform, I believe it is important to challenge the assumptions of free will and basic desert moral responsibility (cf. Fischborn 2018). There is growing evidence that free will beliefs are motivated by the desire to punish others and to justify holding them morally responsible (Clark et al. 2014; Clark, Baumeister, and Ditto 2017; Clark Winegard, and Sharrif 2019). Researchers have also found that free will beliefs correlate with increased punitiveness (Carey and Paulhus 2013; Nadelhoffer and Tocchetto 2013; Shariff et al. 2014). Leaving the concept of free will unchallenged increases the likelihood that our practices and policies will remain focused on individual responsibility and punitive responses to wrongdoing. This, in turn, will stand in the way of adopting the kind of public health alternative I recommend, which shifts public policy, funding, and focus to prevention, rehabilitation, and addressing the social determinants of criminal behavior.

Recent empirical work in social psychology indicates that how we assign responsibility is correlated with prior judgments of what counts as being morally bad, which are in turn dependent upon other larger social and cultural factors (see Hardcastle 2019). Take, for example, psychologist Mark Alicke's *culpable control model* of blame. It proposes that our desire to blame someone intrudes

on our assessments of that person's ability to control his or her thoughts or behavior. As Valerie Hardcastle describes:

> Deciding that someone is responsible for an act, which is taken to be the conclusion of a judgment, is actually part of our psychological process of assessing blame. If we start with a spontaneous negative reaction, then that can lead to our hypothesizing that the source of the action is blameworthy as well as to an active desire to blame that source. This desire, in turn, skews our interpretations of the available evidence such that it supports our blame hypothesis. We highlight evidence that indicates negligence, recklessness, impure motives, or a faulty character, and we ignore evidence that suggests otherwise. In other words, instead of dispassionately judging whether someone is responsible, we validate our spontaneous reaction of blameworthiness. (2018: 320)

Data, in fact, suggest that we often exaggerate a person's actual or potential control over an event to justify our blame judgment and we will even change the threshold of how much control is required for a blame judgment (Neimeth and Sosis 1973; Eften 1974; Sosis 1974; Lerner et al. 1976; Lerner and Miller 1978; Schlenker 1980; Snyder et al. 1983; Alicke 1994, 2000, 2008; Alicke et al. 2008; Alicke Rose and Bloom 2008; Lagnado and Channon 2008; Clark et al. 2014; Everett et al. 2018).

Additional studies by Cory Clark and her colleagues (2014) have shown that a key factor promoting belief in free will is a fundamental desire to blame and hold others morally responsible for their wrongful behaviors. Across five studies they found evidence that greater belief in free will is due to heightened punitive motivations. In one study, an ostensibly real classroom cheating incident led to increased free will beliefs, presumably due to heightened punitive motivations. In a second study, they found that the prevalence of immoral behavior, as measured by crime and homicide rates, predicted free will belief on a country level. Additional studies by Clark, Baumesiter, and Ditto (2017) also demonstrate that free will beliefs are motivated by a desire to punish others and to justify holding them morally responsible. These findings have been replicated and confirmed in meta-analyses (Clark, Winegard, and Shariff 2019). There is good reason to think, then, that our desire to blame and hold others morally responsible comes first and *drives* our belief in free will, rather than the other way around.

Researchers have also found that our judgment on whether an action was done on purpose or not is influenced by our moral evaluation of the outcome of certain actions – that is, whether we morally like or dislike it (Nadelhoffer 2006). And we seem to have an asymmetric understanding of the moral nature of our own actions and those of others, such that we judge our own actions and motivations as more moral than those of the average person (Epley and Dunning 2000). As Maureen Sie describes:

> In cases of other people acting in morally wrong ways we tend to explain those wrongdoings in terms of the agent's lack of virtue or morally bad character traits. We

> focus on those elements that allow us to blame agents for their moral wrongdoings. On the other hand, in cases where we ourselves act in morally reprehensible ways we tend to focus on exceptional elements of our situation, emphasizing the lack of room to do otherwise. (2013: 283)

Additional findings indicate that believing more strongly in free will is correlated with increased punitiveness (Carey and Paulhus 2013; Nadelhoffer and Tocchetto 2013; Shariff et al. 2014), and that weakening free will beliefs, either in general or by offering evidence of an individual's diminished decisional capacity, leads to less punitiveness (Pizarro, Uhlmann, and Salovey 2003; Monterosso, Royzman, and Schwartz 2005; Aspinwall, Brown, and Tabery 2012; Shariff et al. 2014).

For instance, Shariff et al. (2014) hypothesized that if free will beliefs support attributions of moral responsibility, then reducing these beliefs should make people less retributive in their attitudes about punishment. In a series of four studies they tested this prediction and found support for it. In Study 1 they found that people with weaker free will beliefs endorsed less retributive attitudes regarding punishment of criminals, yet their consequentialist attitudes were unaffected. Study 1 therefore supports the hypothesis that free will beliefs positively predict punitive attitudes, and in particular retributive attitudes, yet it also suggests that "the motivation to punish in order to benefit society (consequentialist punishment) may remain intact, even while the need for blame and desire for retribution are forgone" (2014: 7). Study 2 found that experimentally diminishing free will belief through anti-free-will arguments diminished retributive punishment, suggesting a causal relationship (2014: 6). Studies 3 and 4 further found that exposure to stories and findings from neuroscience implying a mechanistic basis for human action similarly produced a reduction in retributivism.

If these empirical findings are any indication, the concepts of free will and basic desert moral responsibility appear intimately connected with increased punitiveness and the desire to blame. Additional work also reveals that belief in free will is correlated with a number of other potentially harmful beliefs, desires, and emotions – including Just World Belief, Right Wing Authoritarianism, and increased Religiosity (Carey and Paulhus 2013; Nadelhoffer and Tocchetto 2013; for overview see Caruso 2019). By abandoning these beliefs, we can look more clearly at the causes and more deeply into the systems that shape individuals and their behavior, and this will allow us to adopt more humane and effective practices and policies. I propose that what we need is a radical paradigm shift in how we view criminal behavior.

It is quite common both in criminal law and everyday attitudes to portray criminal behavior as a failure of moral character and a matter of individual responsible. The retributive justification of legal punishment assumes, for instance, that, absent any excusing conditions, wrongdoers are morally responsible for their actions and *deserve* to be punished in proportion to their bad deeds, even if this provides

no benefits to the individual or society. Since it focuses almost exclusively on the individual and their responsibility, and not on the social determinants of criminal behavior, retributive justice tends to favor punitive approaches to crime rather than policies aimed at targeting the social structures and causes of criminal behavior. The retributive approach maintains that it is the individual who is responsible for criminal wrongdoing, and thus criminal justice is primarily about giving wrongdoers their *just deserts*. Perhaps nobody embodied this ethos of individual responsibility more than Ronald Reagan, who famously said: "We must reject the idea that every time a law's broken, society is guilty rather than the lawbreaker. It is time to restore the American precept that each individual is accountable for his actions."[28]

The problem, however, is that the more we learn about criminal behavior, the more it becomes obvious that crime has more to do with places and circumstances than people. In fact, look closely and you will find that there are lifetimes of trauma, poverty, and social disadvantage that fill the prison system (see Chapter 7). Failing to recognize this has profound consequences. James Dunlea and Larisa Heiphetz (2020), for instance, have studied how adults and children think of people's moral character:

> [They find] that both children and adults often assume that people go to jail or prison because of internal characteristics – that is, something about their personal character led them to the justice system. [Their] research suggests that people very rarely spontaneously think about the social forces, like poverty, housing, and racism, that are enormous contributors to criminal behavior. Instead, they attribute being in jail or prison with internal characteristic ("They're bad people") or behavioral characteristics ("They did a bad thing"). (LaBouff and Dustin 2020)

Importantly, Dunlea and Heiphetz find that these assumptions have consequences. When we think about someone's "badness" as an essential personal quality because we're ignoring the situational forces, we treat them worse and are less generous to them. We adopt punitive reactive responses to crime rather than targeting the situational causes of criminal behavior. We need to radically change our assumptions about criminal behavior and focus instead on those situational forces and changeable behaviors when making policy decisions (LaBouff and Dustin 2020).

Hence, if we seek to reject retributivism and adopt a more holistic and systematic approach to criminal behavior, the belief in free will potentially stands in the way since it encourages punitiveness and is driven by a desire to blame and hold others morally responsible. It sees criminal behavior as primarily a matter of individual responsibility and as a result ends the investigation at precisely the point it should begin. The criminal law, with its assumptions about free will, encourages us to adopt what I call a *time slice approach* to criminal behavior. It asks, at a particular moment in time (the time of the crime), was the agent competent? Were they reasons-

[28] From Reagan's speech at the Republican National Convention, Platform Committee Meeting, July 31, 1968.

responsive? Did they have a guilty mind or criminal intent? Did they understand that their actions were wrong or unlawful? Etc. Etc. If the answers are yes, yes, yes, and yes, then they are legally culpable and it is legitimate to punish them – all things considered and assuming there are no excusing conditions. Of course, the criminal law occasionally considers prior circumstances as relevant (e.g., in cases of domestic violence), but it is primarily focused on establishing *actus reus*, *mens rea*, and the state of mind of the offender at the time of the crime.[29] Unfortunately, adopting such a time slice approach abstracts individuals from their lived circumstances and the social systems they are embedded in. It blinds us to the social determinants of criminal behavior, the causes and systems that shape us, and *how* individuals come to acquire a particular state of mind.

Once we adopt the skeptical perspective, on the other hand, we realize that the myopic focus on individual responsibility, blame, and punishment is mistaken and counterproductive. The skeptical perspective tells us that the lottery of life is not always fair, we do not all have equal starting points, and individuals are embedded in social systems that shape who we are and what we do. In contrast with the time slice approach, it encourages a *historical whole person approach* that sees individuals as byproducts of their histories and circumstances. It helps us to recognize that criminal behavior is often the result of social determinants and that the best way to reduce crime and increase human well-being is to identify and take action on these determinants. So while it is possible to embrace the public health–quarantine model without sharing my skepticism about free will and basic desert moral responsibility, continuing to believe in the latter will work at odds with properly implementing the former. As the empirical findings mentioned earlier indicate, belief in free will, besides being unjustified, is wrapped up in punitive and retributive desires. It keeps alive the notions that wrongdoers justly deserve to suffer for the wrongs they've done and are morally blameworthy in the backward-looking sense. It is time that we leave these antiquated notions behind, lose our moral anger, stop blaming those who find themselves in unfortunate circumstances, and turn our attention to the difficult task of addressing the *causes* that lead to criminality.

[29] Traditionally, juries evaluate two aspects of a case for reasonable doubt: whether the accused actually committed the crime (legally known as *actus reus*) and whether at the moment of the crime the accused understood that he or she was doing wrong (legally known as *mens rea*).