

CHAPTER 11

Hard-Incompatibilist Existentialism

Neuroscience, Punishment, and Meaning in Life

DERK PEREBOOM AND GREGG D. CARUSO

As philosophical and scientific arguments for free will skepticism continue to gain traction, we are likely to see a fundamental shift in the way people think about free will and moral responsibility. Such shifts raise important practical and existential concerns: What if we came to disbelieve in free will? What would this mean for our interpersonal relationships, society, morality, meaning, and the law? What would it do to our standing as human beings? Would it cause nihilism and despair, as some maintain, or would it rather have a humanizing effect on our practices and policies, freeing us from the negative effects of belief in free will? In this chapter, we consider the practical implications of free will skepticism and argue that life without free will and basic desert moral responsibility would not be as destructive as many people believe. We argue that prospects of finding meaning in life or of sustaining good interpersonal relationships, for example, would not be threatened. On treatment of criminals, we argue that although retributivism and severe punishment, such as the death penalty, would be ruled out, preventive detention and rehabilitation programs would still be justified. While we will touch on all these issues herein, our focus will be primarily on this last issue.

We begin in Section 1 by considering two different routes to free will skepticism. The first denies the causal efficacy of the types of willing required for free will and receives its contemporary impetus from pioneering work in neuroscience by Benjamin Libet, Daniel Wegner, and John Dylan Haynes. The second, which is more common in the philosophical literature, does not deny

the causal efficacy of the will but instead claims that whether this causal efficacy is deterministic or indeterministic, it does not achieve the level of control to count as free will by the standards of the historical debate. We argue that while there are compelling objections to the first route (e.g., Al Mele [2009], Eddy Nahmias [2002, 2011], and Neil Levy [2005]), the second route to free will skepticism remains intact. In Section 2, we argue that free will skepticism allows for a workable morality and, rather than negatively impacting our personal relationships and meaning in life, may well improve our well-being and our relationships to others since it would tend to eradicate an often destructive form of moral anger. In Section 3, we argue that free will skepticism allows for adequate ways of responding to criminal behavior—in particular, incapacitation, rehabilitation, and alternation of relevant social conditions—and that these methods are both morally justified and sufficient for good social policy. We present and defend our own preferred model for dealing with dangerous criminals, an incapacitation account built on the right to self-protection analogous to the justification for quarantine (see Pereboom 2001, 2013, 2014a; Caruso 2016a), and we respond to recent objections to it by Michael Corrado and John Lemos.

1. TWO DIFFERENT ROUTES TO FREE WILL SKEPTICISM

In the historical debate, the variety of free will that is of central philosophical and practical importance is the sort required for moral responsibility in a particular but pervasive sense. This sense of moral responsibility is set apart by the notion of *basic desert* (Caruso and Morris 2017; Feinberg 1970; Fischer 2007; Pereboom 2001, 2014a; G. Strawson 1994) and is purely backward-looking and nonconsequentialist. 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 (see Pereboom 2001, 2014a).

Free will skeptics reject this sort of moral responsibility. Rejecting basic desert moral responsibility, however, still leaves other senses intact. For instance, forward-looking accounts of moral responsibility would not be threatened (Pereboom 2014a), nor would the *answerability* sense of moral responsibility defended by Thomas Scanlon (1998) and Hilary Bok (1998). When we encounter apparently immoral behavior, for example, it is perfectly legitimate to ask the agent, “Why did you decide to do that?” or “Do you think it was the right thing to do?” If the reasons given in response to such questions are morally

unsatisfactory, we regard it as justified to invite the agent to evaluate critically what his actions indicate about his intentions and character, to demand apology, or to request reform. Engaging in such interactions is reasonable in light of the right of those harmed or threatened to protect themselves from immoral behavior and its consequences. In addition, we might have a stake in reconciliation with the wrongdoer, and calling him to account in this way can function as a step toward realizing this objective. We also have an interest in his moral formation, and the address described naturally functions as a stage in this process (Pereboom 2012). The thesis of free will skepticism should therefore be understood as the claim 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, not in these other senses.

In the literature, two prominent routes to free will skepticism are identifiable. The first, which is more prominent among scientific skeptics, maintains that recent findings in neuroscience reveal that unconscious brain activity causally initiates action prior to the conscious awareness of the intention to act and that this indicates conscious will is an illusion (e.g., Benjamin Libet, John-Dylan Haynes, Daniel Wegner). The pioneering work in this area was done by Benjamin Libet and his colleagues. In their groundbreaking study on the neuroscience of movement, Libet et al. (1983) investigated the timing of brain processes and compared them to the timing of consciousness will in relation to self-initiated voluntary acts and found that the conscious intention to move (which they labeled *W*) came 200 milliseconds before the motor act, but 350–400 milliseconds after the *readiness potential* (RP)—a ramplike buildup of electrical activity that occurs in the brain and precedes actual movement. Libet and others have interpreted this as showing that the conscious intention or decision to move cannot be the cause of action because it comes too late in the neuropsychological sequence (see Libet 1985, 1999). According to Libet, since we become aware of an intention to act only after the onset of preparatory brain activity, the conscious intention cannot be the true cause of the action (see also Haggard and Eimer 1999; Obhi and Haggard 2004; Pockett 2004; Roediger, Goode, and Zaromb 2008; Soon et al. 2008; Wegner 2002).

Libet's findings, in conjunction with additional findings by John Dylan Haynes (Soon et al. 2008) and others, have led some theorists to conclude that conscious will is an illusion and plays no important causal role in how we act. Haynes and his colleagues, for example, were able to build on Libet's work by using functional magnetic resonance imaging (fMRI) to predict with 60 percent accuracy whether subjects would press a button with either their right or left hand up to ten seconds before the subject became aware of having made that choice (Soon et al. 2008). For some, the findings of Libet and Haynes are enough to threaten our conception of ourselves as free and responsible agents since they appear to undermine the causal efficacy of the types of willing

required for free will. We contend, however, that there are at least three reasons for thinking that these neuroscientific arguments for free will skepticism are unsuccessful.¹

First, there is no direct way to tell which conscious phenomena, if any, correspond to which neural events. In particular, in the Libet studies, it is difficult to determine what the RP corresponds to—for example, is it an intention formation or decision, or is it merely an urge of some sort? Al Mele (2009) has argued that the RP that precedes action by a half-second or more need not be construed as the cause of the action. Instead, it may simply mark the beginning of forming an *intention* to act. According to Mele, “it is much more likely that what emerges around –500 ms is a *potential cause* of a proximal intention or decision than a proximal intention or decision itself” (2009: 51). On this interpretation, the RP is more accurately characterized as an “urge” to act or a preparation to act. That is, it is more accurately characterized as the advent of items in what Mele calls the *preproximal-intention group* (or PPG). We agree with Mele that this leaves open the possibility that conscious intentions can still be causes—that is, if the RP does not correspond to the formation of an intention or decision, but rather an urge, then it remains open that the intention formation or decision is a conscious event.

Second, almost everyone on the contemporary scene who believes we have free will, whether compatibilist or libertarian, also maintains that freely willed actions are caused by virtue of a chain of events that stretch backward in time indefinitely. At some point in time, these events will be such that the agent is not conscious of them. Thus, all free actions are caused, at some point in time, by unconscious events. However, as Eddy Nahmias (2011) correctly points out, the concern for free will raised by Libet’s work is that *all* of the relevant causing of action is (typically) nonconscious, and consciousness is not causally efficacious in producing action. Given determinist compatibilism, however, it is not possible to establish this conclusion by showing that nonconscious events that precede conscious choice causally determine action since such compatibilists hold that every case of action will feature such events and that this is compatible with free will. And, given most incompatibilist libertarianisms, it is also impossible to establish this conclusion by showing that there are nonconscious events that render actions more probable than not by a factor of 10 percent chance (Soon et al. 2008) since almost all such libertarians hold that free will is compatible with such indeterminist causation by unconscious events at some point in the causal chain (De Caro 2011).

Furthermore, Neil Levy raises a related objection when he criticizes Libet’s *impossible demand* (2005) that only consciously initiated actions could be free. Levy correctly argues that this presupposition places a condition upon

1. Some of the criticisms to follow were first made in Bjornsson and Pereboom (2014).

freedom of action which is in principle impossible to fill for reasons that are entirely conceptual and have nothing to do, per se, with Libet's empirical findings. As Levy notes, "Exercising this kind of control would require that we control our control system, which would simply cause the same problem to arise at a higher-level or initiate an infinite regress of controllings" (2005: 67). If the unconscious initiation of actions is incompatible with control over them, then free will is impossible on conceptual grounds. Thus, Libet's experiments do not constitute a separate, empirical challenge to our freedom (see Levy 2005).

Finally, several critics have correctly noted the unusual nature of the Libet-style experimental situation—that is, one in which a conscious intention to flex at some time in the near future is already in place, and what is tested for is the specific implementation of this general decision. Nahmias (2011), for example, convincingly points out that it is often the case—when, for instance, we drive or play sports or cook meals—that we form a conscious intention to perform an action of a general sort, and subsequent specific implementation are not preceded by more specific conscious intentions. But, in such cases, the general conscious intention is very plausibly playing a key causal role. In Libet-style situations, when the instructions are given, subjects form conscious intentions to flex at some time or other, and, if it turns out that the specific implementations of these general intentions are not in fact preceded by specific conscious intentions, this would be just like the kinds of driving and cooking cases Nahmias cites. It seems that these objections cast serious doubts on the potential for neuroscientific studies to undermine the claim that we have the sort of free will at issue.

Before moving on to the second route to free will skepticism, it is worth quickly noting that there are other scientific threats to free will in addition to those posed by neuroscience. Recent work in psychology and social psychology on *automaticity*, *situationism*, and the *adaptive unconscious*, for instance, has shown that the causes that move us are often less transparent to ourselves than we might assume—diverging in many cases from the conscious reasons we provide to explain and/or justify our actions (e.g., Bargh 1997, 2008; Bargh and Chartrand 1999; Bargh and Ferguson 2000; Doris 2002; Nisbett and Wilson 1977; Wilson 2002). These findings reveal just how wide open our internal psychological processes are to the influence of external stimuli and events in our immediate environment, without knowledge or awareness of such influence. They also reveal the extent to which our decisions and behaviors are driven by implicit biases (see Greenwald, McGhee, and Schwartz 1998; Kang et al. 2010; Nosek et al. 2007; Uhlmann and Cohen 2005). No longer is it believed that only "lower level" or "dumb" processes can be carried out nonconsciously. We now know that the higher mental processes that have traditionally served as quintessential examples of "free will"—such as evaluation and judgment, reasoning and problem-solving, and interpersonal

behavior—can and often do occur in the absence of conscious choice and guidance (Bargh and Ferguson 2000: 926; see also Wilson 2002).

While these findings would not be enough, on their own, to establish global skepticism about free will and basic desert moral responsibility (see Levy 2014), they represent a potential threat to our everyday folk understanding of ourselves as conscious, rational, responsible agents since they indicate that the conscious mind exercises less control over our behavior than we have traditionally assumed. Even some compatibilists now admit that because of these findings “free will is at best an occasional phenomenon” (Baumeister 2008: 17; see also Nahmias). This is an important concession because it acknowledges that the *threat of shrinking agency*—as Thomas Nadelhoffer (2011) calls it—remains a serious one independent of the neuroscientific concerns just discussed. The deflationary view of consciousness which emerges from these empirical findings, including the fact that we often lack transparent awareness of our true motivational states, is potentially agency-undermining and could shrink the realm of morally responsible action (see Caruso 2012, 2015, 2016*b*; King and Carruthers 2012; Levy 2014; Nadelhoffer 2011; Sie and Wouters 2010). It is important therefore that accounts of moral responsibility that require, for instance, reasons-responsiveness or evaluation of personal-level attitudes (including beliefs, commitments, and goals) make explicit the role they see consciousness playing and the extent to which automaticity, situationism, and implicit bias may limit or restrict morally responsible behavior. For our purposes, however, we are going to table these concerns for the remainder of this chapter to focus on a second route to free will skepticism—one, which we maintain, is more successful at establishing a global skepticism about free will and basic desert moral responsibility.

In the past, the standard argument for free will skepticism was based on the notion of *determinism*—the thesis that every event or action, including human action, is the inevitable result of preceding events and actions and the laws of nature. *Hard determinists* argued that determinism is true and incompatible with free will and basic desert moral responsibility, either because it precludes the *ability to do otherwise* (leeway incompatibilism) or because it is inconsistent with one’s being the “ultimate source” of action (source incompatibilism). While hard determinism had its classic statement in the time when Newtonian physics reigned, it has very few defenders today—largely because the standard interpretation of quantum mechanics has been taken by many to undermine, or at least throw into doubt, the thesis of universal determinism. We nonetheless maintain that even if you allow some indeterminacy to exist at the microlevel of our existence—the level studied by quantum mechanics—the sort of free will at issue in the historical debate would still be threatened. Our view differs from hard determinism, then, in that it maintains that, whatever the fundamental nature of reality, we would still lack free will. A more accurate name for our position would therefore be *harm*

[198] *Derk Pereboom and Gregg D. Caruso*

incompatibilism (see Pereboom 2001, 2014*a*), to differentiate it from hard determinism. Hard incompatibilism does not deny the causal efficacy of the will but instead claims that whether this causal efficacy is deterministic or indeterministic, it does not achieve the level of control required for basic desert moral responsibility.

Hard incompatibilism amounts to a rejection of both compatibilism and libertarianism. It maintains that the sort of free will required for basic desert moral responsibility is incompatible with causal determination by factors beyond the agent's control and *also* with the kind of indeterminacy in action required by the most plausible versions of libertarianism. Against the view that free will is compatible with the causal determination of our actions by natural factors beyond our control, we argue that there is no relevant difference between this prospect and our actions being causally determined by manipulators (see Pereboom 2001, 2014*a*). Against event causal libertarianism, we advance the disappearing agent objection, according to which agents are left unable to settle whether a decision occurs and hence cannot have the control required for moral responsibility (Caruso 2012; Pereboom 2001, 2014*a*). The same problem, we contend, arises for noncausal libertarian accounts, which also fail to provide agents with the control in action required for basic desert moral responsibility. While agent-causal libertarianism could, in theory, supply this sort of control, we argue that it cannot be reconciled with our best physical theories (Pereboom 2001, 2014*a*) and faces additional problems accounting for mental causation (Caruso 2012). Since this exhausts the options for views on which we have the sort of free will at issue, we conclude that free will skepticism is the only remaining position.

Since the arguments for hard incompatibilism have been spelled out and defended at great length elsewhere (see, e.g., Caruso 2012, 2014; Pereboom 2001, 2014*a*, 2014*b*), and no solid refutations of them have yet been offered (cf. Fischer 2014; Nelkin 2014; for a reply Pereboom 2014*b*), we will not elaborate on them further here. Instead, we will now shift our attention to exploring the practical implications of free will skepticism. For many, it is not the philosophical arguments for free will skepticism that are the problem, it is the existential angst they create and the fear that relinquishing belief in free will and basic desert moral responsibility would undermine morality, negatively affect our interpersonal relationships, and leave us unable to adequately deal with criminal behavior. To these concerns we now turn.

2. MORAL AND PERSONAL IMPLICATIONS

If the argument for free will skepticism is convincing, one can conclude that we lack the sort of free will required for moral responsibility in the basic desert sense. The concern for the skeptical position is not that there is considerable

empirical evidence that it is false or that there is a challenging argument for its incoherence. The main question it raises is instead practical: Can we live with the belief that it is true? A number of free will skeptics, including Honderich (1988), Pereboom (1995, 2001, 2014a), Levy (2011), and Caruso (2012, 2017a, forthcoming) argue that we, in fact, can. We will begin by briefly addressing two practical issues. The first concerns the extent to which the skeptic can retain our ordinary conception of morality and responsibility, the second the degree to which it coheres with the emotions required for the kinds of personal relationships we value. We will then discuss at length the implications of the view for treatment of criminals.

2.1. Free Will Skepticism, Morality, and Responsibility

Accepting free will skepticism requires rejecting our ordinary view of ourselves as blameworthy or praiseworthy in the basic desert sense. A critic might first object that if we gave up this belief, we could no longer count actions as morally bad or good. In response, even if we came to hold that a serial killer was not blameworthy due to a degenerative brain disease, we could still justifiably agree that his actions are morally bad. Still, secondly, the critic might ask, if determinism precluded basic desert blameworthiness, would it not also undercut judgments of moral obligation? If “ought” implies “can,” and if because determinism is true an agent could not have avoided acting badly, it would be false that she ought to have acted otherwise. Furthermore, if an action is wrong for an agent just in case she is morally obligated not to perform it, determinism would also undermine judgments of moral wrongness (Haji 1998). In response, we contend that even if the skeptic were to accept all of this (and she might resist at various points; cf. Pereboom 2014a: ch. 6; Waller 2011), axiological judgments of moral goodness and badness would not be affected (Haji 1998; Pereboom 2001). So, in general, free will skepticism can accommodate judgments of moral goodness and badness, which are arguably sufficient for moral practice.

Third, the critic might object that if we stopped considering agents as blameworthy in the basic desert sense, we would be left with insufficient resources for addressing immoral behavior (e.g., Nichols 2007). However, the skeptic might turn instead to other senses of moral responsibility that have not been a focus of the free will debate (Pereboom 2013; 2014a: ch. 6). Our moral practice features a number of senses of moral responsibility, some of which do not invoke basic desert. For instance, when we encounter immoral action, we might ask the agent to consider what his actions indicate about his intentions and character, to demand apology, or to request reform, thereby having him consider reasons to behave differently in the future. Engaging in such interactions counts as reasonable in view of the

[200] *Derk Pereboom and Gregg D. Caruso*

right of those wronged or threatened by wrongdoing to protect themselves from bad behavior and its consequences. Our practice also features an interest in the wrongdoer's moral formation, and the address described naturally functions as a step in this process. Moreover, our practice also has a stake in our reconciliation with the wrongdoer, and calling him to account plausibly serves as a stage in securing this aim. Such interactions, because they address the agent's capacity to consider and respond to reasons, manifest respect for her as a rational agent. The main thread of the historical free will debate does not pose causal determination as a challenge to this sense of moral responsibility, and thus this is an aspect of our practice that the free will skeptic can endorse.

2.2. Personal Relationships and Meaning in Life

Is the assumption that we are morally responsible in the basic desert sense required for the sorts of personal relationships we value? The considerations raised by P. F. Strawson in his essay "Freedom and Resentment" (1962) suggest a positive answer. In his view, 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 practice is grounded in this way, the truth or falsity of causal determinism is not relevant to whether we justifiably hold each other and ourselves morally responsible. Moreover, if causal determinism were true and did threaten these attitudes, as the free will skeptic is apt to maintain, we would face instead the prospect of the cold and calculating objectivity of attitude, a stance that relinquishes the reactive attitudes. In Strawson's view, adopting this stance would rule out the possibility of the meaningful sorts of personal relationships we value.

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, 2014a). 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.

Guilt is also imperiled by free will skepticism, and this consequence would seem to be more difficult to accommodate. The skeptic's view stands to undercut guilt because it would seem to involve the supposition that one is blameworthy in the basic desert sense for an immoral action one has performed. There is much at stake here, the critic might contend, because, absent guilt, we would not be motivated to moral improvement after acting badly, and we would be kept from reconciliation in impaired relationships. In addition, the critic continues, because guilt is undermined by the skeptical view, repentance is also ruled out because feeling guilty is a prerequisite for a repentant attitude. In response, suppose instead you acknowledge that you have acted immorally, and, as Bruce Waller advocates, you feel deep sorrow for what you have done (Waller 1990: 165–166; cf. Bok 1998); as a result, you are motivated to eradicate your disposition to behave in this bad way. This response can secure the good that guilt can also secure, and it is wholly compatible with the free will skeptic's view.

Gratitude arguably presupposes that the person to whom one is grateful is praiseworthy in the basic desert sense for a beneficial act (cf. Honderich 1988: 518–519). But even if this is so, certain aspects of gratitude would not be undercut, and these aspects would seem to provide what is required for the personal relationships we value. Gratitude involves being thankful toward the person who has acted beneficially. This aspect of gratitude is in the clear; one can be thankful to a young child for some kindness without supposing that she is praiseworthy in the basic desert sense. Gratitude typically also involves joy as a response to what someone has done, and free will skepticism does not yield a challenge to being joyful and expressing joy when others act beneficially.

Perhaps some of the recommended transformations in emotional attitudes may not be possible for us. For example, in certain situations, refraining from moral anger may be beyond our power, and thus even the committed skeptic might not be able to make the change the skeptical view suggests. At this point Shaun Nichols (2007) invokes the distinction between *narrow-profile* emotional responses—local or immediate emotional reactions to situations—and *wide-profile* responses, which are not immediate and involve rational reflection. We might expect to be unable to appreciably reduce narrow-profile moral anger as an immediate reaction upon being deeply hurt in an intimate personal relationship. In wide-profile cases, however, diminishing or even eliminating moral anger is open, or at least disavowing it in the sense of rejecting any force it may be assumed to have in justifying a harmful response to wrongdoing. This modification of moral anger might well be advantageous for our valuable personal relationships, and it stands to bring about the equanimity that Spinoza thought free will skepticism, more generally, would secure.

3. FREE WILL SKEPTICISM AND CRIMINAL BEHAVIOR

One of the most frequently voiced criticisms of free will skepticism is that it is unable to adequately deal with criminal behavior and that the responses it would permit as justified are insufficient for acceptable social policy. This concern is fueled by two factors. The first is that one of the most prominent justifications for punishing criminals, retributivism, is incompatible with free will skepticism. The second is that alternative justifications that are not ruled out by the skeptical view *per se* face significant independent moral objections. Yet, despite this concern, we maintain that free will skepticism leaves intact other ways to respond to criminal behavior—in particular incapacitation, rehabilitation, and alteration of relevant social conditions—and that these methods are both morally justifiable and sufficient for good social policy. In this section, we present and defend our preferred model for dealing with dangerous criminals, an incapacitation account built on the right to self-protection analogous to the justification for quarantine (see Caruso 2016*a*, 2017*b*; Pereboom 2001, 2013, 2014*a*), and respond to objections to it by John Lemos (2016) and Michael Corrado (2016).

To begin, we need to recognize that retributive punishment is incompatible with free will skepticism because it maintains that punishment of a wrongdoer is justified for the reason that he *deserves* something bad to happen to him just because he has knowingly done wrong—this could include pain, deprivation, or death. As Douglas Husak puts it, “Punishment is justified only when and to the extent it is deserved” (2000: 82). And Mitchell Berman writes, “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). Furthermore, for the retributivist, it is the *basic* desert attached to the criminal’s immoral action alone that provides the justification for punishment. The desert the retributivist invokes is basic in the sense that justifications for punishment that appeal to it are not reducible to consequentialist considerations nor to goods such as the safety of society or the moral improvement of the criminal.

Free will skepticism undermines this justification for punishment because it does away with the idea of basic desert. If agents do not deserve blame just because they have knowingly done wrong, neither do they deserve punishment just because they have knowingly done wrong. The challenge facing free will skepticism, then, is to explain how we can adequately deal with criminal behavior without the justification provided by retributivism and basic desert. While some critics contend this cannot be done, free will skeptics point out that there are several alternative ways of justifying criminal punishment (and dealing with criminal behavior more generally) that do not appeal to the notion of basic desert and are thus not threatened by free will skepticism.

These include moral education theories, deterrence theories, punishment justified by the right to harm in self-defense, and incapacitation theories. While we maintain the first two approaches face independent moral objections—objections that, though perhaps not devastating, make them less desirable than their alternative—we argue that an incapacitation account built on the right to harm in self-defense provides the best option for justifying a policy for treatment of criminals consistent with free will skepticism. Before turning to our positive account, let us briefly say something about the first two alternative approaches.

Moral education theories draw an analogy with justification of the punishment of children. Children are typically not punished to exact retribution, but rather to educate them morally. Since moral education is a generally acceptable goal, a justification for criminal punishment based on this analogy is one the free will skeptic can potentially accept. Despite its consistency with free will skepticism, though, a serious concern for this type of theory is that it is far from evident that punishing adult criminals is similarly likely to result in moral improvement. Children and adult criminals differ in significant respects. For example, adult criminals, unlike children, typically understand the moral code accepted in their society. Furthermore, children are generally more psychologically malleable than are adult criminals. For these and other reasons, we see this approach as less desirable than an alternative incapacitation account (see Pereboom 2014a: ch. 7).

Deterrence theories, especially utilitarian deterrence theories, have probably been the most discussed alternative to retributivism. According to deterrence theories, the prevention of criminal wrongdoing serves as the good on the basis of which punishment is justified. The classic deterrence theory is Jeremy Bentham's. In his conception, the state's policy on criminal behavior should aim at maximizing utility, and punishment is legitimately administered if and only if it does so. The pain or unhappiness produced by punishment results from the restriction on freedom that ensues from the threat of punishment, the anticipation of punishment by the person who has been sentenced, the pain of actual punishment, and the sympathetic pain felt by others such as the friends and family of the criminal (Bentham 1823/1948). The most significant pleasure or happiness that results from punishment derives from the security of those who benefit from its capacity to deter.

While deterrence theories are completely compatible with free will skepticism, there are three general moral objections against them. The first is that they will justify punishments that are intuitively too severe. For example, it would seem that, in certain cases, harsh punishment would be a more effective deterrent than milder forms, while the harsh punishments are intuitively too severe to be fair. The second concern is that such accounts would seem to justify punishing the innocent. If, for instance, after a series of horrible crimes the actual perpetrator is not caught, potential criminals might come to believe

that they can get away with serious wrongdoing. Under such circumstances it might maximize utility to frame and punish an innocent person. Last, there is the “use” objection, which is a problem for utilitarianism more generally. Utilitarianism sometimes requires people to be harmed severely, without their consent, in order to benefit others, and this is often intuitively wrong. While some skeptics believe these objections can be met, we recommend that free will skeptics seek a different alternative to retributivism.

There is, however, a legitimate theory for prevention of especially dangerous crime that is neither undercut by free will skepticism nor by other moral considerations. This theory is based on an analogy with quarantine and draws on a comparison between treatment of dangerous criminals and treatment of carriers of dangerous diseases. The free will skeptic claims that criminals are not morally responsible for their actions in the basic desert sense. Plainly, many carriers of dangerous diseases are not responsible in this or in any other sense for having contracted these diseases. 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. For similar justificatory reasons, we argue, 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.

One might justify both quarantine in the case of disease and incapacitation of dangerous criminals on purely utilitarian or consequentialist grounds. But we want to resist this strategy. Instead, on our view, incapacitation of the 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 objection, as will become clear in what follows.

It is important to see that this analogy places several constraints on the treatment of criminals. First, as less dangerous diseases justify only preventative measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate restraints. In fact, for certain minor crimes perhaps only some degree of monitoring could be defended. Second, 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 (cf. D’Angelo 1968: 56–59). Third, 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. Finally, there are measures for preventing crime more generally, such as providing for

adequate education and mental health care, which the free will skeptic can readily endorse.

We contend that this account provides a more resilient proposal for justifying criminal sanctions than either the moral education or deterrence theories. One advantage this approach has over the utilitarian deterrence theory 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 (Pereboom 2001, 2013, 2014a). Our account therefore maintains 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. Furthermore, the less dangerous the disease, the less invasive the justified prevention methods would be, and similarly, the less dangerous the criminal, the less invasive the justified forms of incapacitation would be.

In addition to these restrictions on harsh and unnecessary treatment, our account also advocates for a broader approach to criminal behavior that moves beyond the narrow focus on sanctions. Consider, for example, the recent proposal by Caruso (2016a, 2017b) to place the quarantine analogy within the broad 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. If we care about public health and safety, the focus should always be on preventing crime from occurring in the first place by addressing the systemic causes of crime. Prevention is always preferable to incapacitation.

Furthermore, public health ethics sees *social justice* as a foundational cornerstone to public health and safety (Caruso 2017b). 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

[206] Derk Pereboom and Gregg D. Caruso

of the criminal justice system is to identify and remedy social and economic inequalities responsible for crime. Just as public health is negatively affected by poverty, racism, and systematic inequality, so, too, is public safety. This broader approach to criminal justice therefore places issues of social justice at the forefront. It sees racism, sexism, poverty, and systemic disadvantage as serious threats to public safety, and it prioritizes the reduction of such inequalities (see Caruso 2017b).

Summarizing our account, then, the core idea is that the right to harm in self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. The resulting account would not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. Our account also specifies attention to the well-being of criminals, which would change much of current policy. Furthermore, free will skeptics would continue to endorse measures for reducing crime that aim at altering social conditions, such as improving education, increasing opportunities for fulfilling employment, and enhancing care for the mentally ill. This combined approach to dealing with criminal behavior, we argue, is sufficient for dealing with dangerous criminals, leads to a more humane and effective social policy, and is actually preferable to the harsh and often excessive forms of punishment that typically come with retributivism.

Michael Corrado raises three objections to this incapacitation account, which leads him to reject it in favor of a compromise view, which he calls Correction. This position, while denying basic desert moral responsibility, endorses hard treatment of reasons-responsive criminals on the ground of moral educational benefit to the criminal and deterrence of future crime. Corrado's first objection is that our view, unlike his, makes no distinction between people who are dangerous and yet have the sort of control captured by the reasons-responsiveness condition, and those who are dangerous but lack this sort of control, and instead treats all criminals on the model of illness. The second is that, given our view, too many people will be drawn into the criminal justice system since merely posing a danger is sufficient to make one a candidate for incapacitation. The third objection is that those who are incapacitated would need to be compensated, and this would be prohibitively costly.

On the first concern, in *Living Without Free Will*, Pereboom distinguished his position from views according to which criminal tendencies are exclusively psychological illnesses, modeled on physical illness (Pereboom 2001). It is true that on our view policies for making a detained criminal safe for release would address a condition in the offender that results in the criminal behavior. But such conditions are not restricted to psychological illnesses; they also include conditions that are not plausibly classified as illness, such as insufficient sympathy for others or a strong tendency to assign blame to others and not to

oneself when something goes wrong. What unites policies for treatment of criminals on our view is not that they assume that they are psychologically ill and therefore in need of psychiatric treatment. Instead, they all aim to bring about moral change in an offender by nonpunitively addressing conditions that underlie criminal behavior.

What sets the illness model apart is that proposed treatment does not address the criminal's capacity to respond to reasons, but circumvents such capacities. For example, consider the Ludovico method, made famous by Anthony Burgess's book and Stanley Kubrick's film *A Clockwork Orange*. Alex, a violent criminal, is injected with a drug that makes him nauseous while at the same time he is made to watch films depicting the kind of violence to which he is disposed. The goal of the method is that the violent behavior be eliminated by generating an association between violence and nausea. Herbert Morris's objection to therapy of this sort is that the criminal is not changed by being presented with reasons for altering his behavior which he would autonomously and rationally accept. But Pereboom (2001) cites a number of programs for treating criminals that are not in accord with the illness model. The Oregon Learning Center, for instance, aims to train parents and families to formulate clear rules, monitor behavior, and to set out fair and consistent procedures for establishing positive and negative incentives. The method involves presentation of reasons for acting and strategies for realizing aims in accord with these reasons. This program is successful: in one study, youth in ten families showed reductions of 60 percent in aggressive behavior compared to a 15 percent drop in untreated control families.²

Pereboom also cites therapeutic programs designed to address problems for the offender's cognitive functioning. A number of cognitive therapy programs are inspired by S. Yochelson and S. Samenow's influential work *The Criminal Personality* (1976), which argues that certain kinds of cognitive distortions generate and sustain criminal behavior. Kris Henning and Christopher Frueh provide some examples of such cognitive distortions:

Car thieves would be more likely to continue with their antisocial activities if they reasoned that *stealing cars isn't as bad as robbing people* (minimization of offense) or *I deserve to make a couple of bucks after all the cops put me through last time (taking the role of the victim)*. Similarly, a rapist who convinces himself, *she shouldn't have been wearing that dress if she didn't want me to touch her (denial of responsibility)*, would probably be at greater risk to reoffend than someone who accepts responsibility for his actions. (1996: 525)

2. Patterson, Chamberlain, and Reid (1982), cited in Walters (1992: 143). Cf. Patterson (1982), Alexander and Parsons (1982). For a review of studies on family therapy, see Gendreau and Ross (1979).

In 1988, the state of Vermont put in place a therapeutic program inspired by the Yochelson and Samenow's cognitive distortion model. The Cognitive Self-Change Program was initially designed as group treatment for imprisoned male offenders with a history of interpersonal aggression, and it later included imprisoned nonviolent offenders. Henning and Frueh provide a description of the procedure:

Treatment groups met 3–5 times per week. During each session, a single offender was identified to present a “thinking report” to the group. Typically, these reports documented prior incidents of anti-social behavior, although more current incidents were reported on when appropriate. At the beginning of each session, the offender would provide the group with an objective description of the incident. He would then list all of the thoughts and feelings he had before, during, and after the event. After the report was delivered, the group worked with the offender to identify the cognitive distortions that may have precipitated the antisocial response to the situation. Role plays sometimes were used during these sessions to develop a better understanding of the cognitions and emotions that led up to the offender's behavior. Once an offender learned to identify his primary criminogenic thought patterns, intervention strategies were discussed in the group to help him prevent such distortions from occurring in the future. These might include cognitive strategies (e.g. challenging one's cognitions, cognitive redirection) and/or behavioral interventions (e.g. avoidance of high-risk situations; discussion of cognitions and feelings with therapist, friend, or partner). (1996: 525)

Henning and Frueh found that in a group of 28 who had participated in this program, 50 percent (14) were charged with a new crime following their release. In a control group of 96 who had not participated, 70.8 percent (68) were charged with a new offense. Twenty-five percent of offenders who had participated received a new criminal charge within one year, 38 percent within two years, and 46 percent within three. By contrast, in the comparison group, 46 percent had been charged with a new crime within one year, 67 percent within two, and 75 percent within three. These results were found to be statistically significant.

Models of *restorative justice* proved another alternative for rehabilitating criminals in a way that respects the reasons-responsiveness of agents. It also has the additional benefit of addressing the rights of victims by having the criminal admit the wrong done, acknowledge the harm caused, and agree to work toward reconciliation with the victim or the victim's families. Models of restorative justice are perfectly consistent with free will skepticism as long as they are employed in a way that does not appeal to backward-looking blame in the restorative process. Consider, for instance, the recent success of schools in using restorative methods as an alternative

to school suspension.³ In traditional school discipline programs, students face an escalating scale of punishment for infractions that can ultimately lead to expulsion. There is now strong research, however, that shows pulling students out of class as punishment can hurt their long-term academic prospects (Losen et al. 2015; Losen, Martinez, and Okelola 2014; Richmond 2015). Furthermore, data show that punishments are often distributed unequally. More black students, for example, are suspended nationally than white students (Richmond 2015; US Department of Education Office for Civil Rights 2014).

As an alternative, public schools from Maine to Oregon have begun to employ restorative justice programs designed to keep students in school while addressing infractions in a way that benefits both the offender and the offended. Here is one description of such a program:

Lower-level offenses can be redirected to the justice committee, which is made up of student mediators, with school administrators and teachers serving as advisors. The goal is to provide a nonconfrontational forum for students to talk through their problems, addressing their underlying reasons for their own behaviors, and make amends both to individuals who have been affected as well as to the larger school community. (Richmond 2015)

Students are often given the option of participating in these alternative programs or accepting traditional discipline, including suspension. As reported on in *The Atlantic*, “Early adopters of the practice report dramatic declines in school-discipline problems, as well as improved climates on campuses and even gains in student achievement” (Richmond 2015). Programs like this reveal that the more punitive option—for example, expulsion rather than restorative processes—is often less effective from the perspective of future protection, future reconciliation, and future moral formation. They also reveal how rehabilitating individuals can be done in a way that appeals directly to their reasons-responsive capacities.

Contrary to Corrado’s concerns, then, we maintain that methods of therapy that engage reasons-responsive abilities should be preferred. On the forward-looking account of moral responsibility we endorse (Pereboom 2013; 2014a: ch. 6), when we call an agent to account for immoral behavior, at the stage of moral address we request an explanation with the intent of having the agent acknowledge a disposition to act badly, and then, if she has in fact so acted without excuse or justification, we aim for her to come to see that the disposition issuing in the action is best eliminated. In normal cases, this

3. See “Alternative to School Suspension Explored Through Restorative Justice” (Associated Press), December 17, 2014; “When Restorative Justice in Schools Works” (*The Atlantic*), December 29, 2015 (Richmond 2015).

change is produced by way of the agent's recognition of moral reasons to eliminate the disposition. Accordingly, it is an agent's responsiveness to reasons—together with the fact that we have a moral interest in our protection, her moral formation, and our reconciliation with her—that explains why she is an appropriate recipient of blame in this forward-looking sense. While many compatibilists see some type of attunement to reasons as the key condition for basic desert moral responsibility, we instead view it as the most significant condition for a notion of responsibility that focuses on future protection, future reconciliation, and future moral formation.

Still, a concern for many forms of therapy proposed for altering criminal tendencies is that they circumvent, rather than address, the criminal's capacity to respond to reasons. On our view, forms of treatment that do address reasons-responsiveness are to be preferred. However, the fact that a mode of therapy circumvents rather than addresses the capacities that confer dignity on us should not all by itself make it illegitimate for agents who are in general responsive to reasons but not in particular respects. Imagine such an agent who is beset by bouts of violent anger that he cannot control in some pertinent sense. Certain studies suggest that this tendency is due to deficiencies in serotonin and that it can sometimes be alleviated by antidepressants.⁴ It would seem mistaken to claim that such a mode of treatment is illegitimate because it circumvents capacities for rational and autonomous response. In fact, this sort of treatment often produces responsiveness to reasons where it was previously absent (Pereboom 2001). A person beset by violent anger will typically not be responsive to certain kinds of reasons to which he would be responsive if he were not suffering from this problem. Therapy of this sort can thus increase reasons-responsiveness. By analogy, one standard form of treatment for alcoholism—which many alcoholics voluntarily undergo—involves the use of a drug, Antabuse, which makes one violently ill after the ingestion of alcohol. By counteracting addictive alcoholism, this drug can result in enhanced reasons-responsiveness.

Furthermore, suppose that despite serious attempts at moral rehabilitation that do not circumvent the criminal's rational capacities, and despite procedures that mechanically increase the agent's capacities for reasons-responsiveness, the criminal still displays dangerously violent tendencies. Imagine that the choice is now between indefinite confinement without hope for release and behavioristic therapy that does not increase the agent's capacity for reasons-responsiveness. It is not obvious that here the behavioristic therapy should be ruled out as morally illegitimate. One must assess the appropriateness of therapy of this kind by comparing it with the other options. Suppose, for example, that the only legitimate alternative to confinement for life is application of some behavioristic therapy. It is not clear that,

4. Burlington Free Press (Associated Press), December 15, 1997, p. 1.

under such circumstances, the moral problems with such a therapy are not outweighed—especially if it is carried out in a way that respects autonomy by leaving the decision up to the criminal.

Behavioristic therapies, however, are almost always suboptimal when compared with their alternatives—that is, methods that directly appeal to a criminal’s rational capacities or, when these fail, therapies that mechanically increase the agent’s capacities for reasons-responsiveness. There are also additional alternatives to behavioral therapy that, at least in the future, may prove more successful in rehabilitating criminals. The use of neurofeedback, for instance, in correctional settings has been suggested as “an innovative approach that may ultimately lessen criminal behavior, prevent violence, and lower recidivism” (Gkotsi and Benaroyo 2012: 3; see also Evans 2006; Quirk 1995; Smith and Sams 2005). As Gkotsi and Benaroyo describe:

Neurofeedback or neurotherapy is a relatively new, noninvasive method which is based on the possibility of training and adjusting the speed of brainwaves, which normally occur at various frequencies (Hammond, 2011). An overabundance, or deficiency in one of these frequencies, often correlates with conditions such as depression, and emotional disturbances and learning disabilities, such as Attention Deficit Hyperactivity Disorder (ADHD) (Greteman, 2009). . . . Therapists attach electrodes to the patients’ head and a device records electrical impulses in the brain. These impulses are sorted into different types of brain waves. Using a program similar to a computer game, patients learn to control the video display by achieving the mental state that produces increases in the desired brain wave activity. Neurofeedback has gained recognition for its potential benefits for children with ADHD, alcoholics and drug addicts. It can also enhance athlete and musician performance as well as improve elderly people’s cognitive function (Greteman 2009)

Douglas Quirk, a Canadian researcher, tested the effects of a neurofeedback treatment program on seventy-seven dangerous offenders in an Ontario correctional institute who suffered from deep-brain epileptic activity. The results demonstrated reduction in the subjects’ criminal recidivism and suggested that “a subgroup of dangerous offenders can be identified, understood and successfully treated using this kind of biofeedback conditioning program” (Quirk 1995; as quoted by Gkotsi and Benaroyo 2012: 3). Additional studies by Smith and Sams (2005) on juvenile offenders with significant psychopathology and electroencephalographic abnormalities, and by Martin and Johnson (2005) on male adolescents diagnosed with ADHD also demonstrated reduced recidivism, improved cognitive performance, improved emotional and behavioral reactions, and inhibition of inappropriate responses.

More invasive than neurofeedback is another potential treatment: deep brain stimulation (DBS). DBS has been used as a last-resort treatment of

[212] *Derk Pereboom and Gregg D. Caruso*

neuropsychological disorders including schizophrenia, Parkinson's disease, dystonia, Tourette's syndrome, pain, depression, and obsessive compulsive disorder. It involves the surgical placement of a device in the brain that sends electrical impulses to target areas that have been linked to the particular condition. Some neurologists and neuroscientists have recently proposed that DBS can be used for the rehabilitation of criminal psychopaths (Center for Science and Law 2012; Hoeprich 2011).

Since there are very few options currently available for the effective rehabilitation of psychopaths, which often leaves continued incapacitation as society's sole means to protection, some have argued that DBS may provide a better and more effective alternative (see Hoeprich 2011). As the Center for Science and Law describe:

Psychopaths have been shown to have neurophysiological deficiencies in various brain structures compared to healthy human subjects. These structures include the amygdala (an important center for processing of emotionally-charged and stimulus-reward situations) and the ventromedial prefrontal cortex (suppression of emotional reactions and decision-making). DBS can potentially be used, then, in these areas to see if psychopathic tendencies can be suppressed. (2012)

There are, however, important ethical concerns with regard to the use of DBS for rehabilitating psychopaths (see Gkotsi and Benaroyo 2012), especially since it is highly experimental, with many reported negative side effects, and is far more invasive than neurofeedback, which is generally believed to be a fairly safe procedure.

We propose, then, that rehabilitation methods that directly appeal to a criminal's rational capacities should always be preferred and attempted first. When these fail, we contend that it is sometimes acceptable to employ therapies that mechanically increase an agent's capacities for reasons-responsiveness but that these therapies should involve the participation of the subject to the greatest extent possible (e.g., talk therapies in conjunction with other forms of treatment), should involve the consent of the subject, and should be ordered such that noninvasive methods are prioritized. When all else fails and only more invasive methods are left—for example, DBS for psychopaths—important ethical questions need to be considered and answers weighed, but leaving the final choice up to the subject is an attractive option.

3.1. The Scope Issue

Corrado's second objection is that too many people will be drawn into the criminal justice system on our account. First, Corrado intimates that many

more people would be detained than is the case currently. Second, there is the issue of incapacitating those who pose threats but have not yet committed a crime. Corrado is reasonably concerned about the prospects of such a policy.

On the first issue, in all of our writings on this topic we have in effect advocated the principle of least infringement, which specifies that the least restrictive measures should be taken to protect public health and safety. While we do believe that we should indefinitely detain mass murderers and serial rapists who cannot be rehabilitated and remain threats, we do not believe that non-violent shoplifters who remain threats and cannot be rehabilitated should be preventatively detained at all, by contrast with being monitored, for example. Our view does not prescribe that all dangerous people be detained until they are no longer dangerous. Certain kinds of persisting threats can be dealt with by monitoring, in contrast with detention. Moreover, other behavior that is currently considered criminal might not require incapacitation at all. Our view is consistent, for example, with the decriminalization of nonviolent behavior such as recreational drug use and thus is consistent with many fewer people being detained than in the United States currently.

In addition to monitoring and decriminalization, monetary fines could also serve as suitable sanctions for low-level crimes. When someone fails to heed a stop sign, for example, they put at risk the potential safety of others. The right of self-protection and the prevention of harm to others justify liberty-limiting laws backed by the threat of sanctions, but the sanctions in this case would need to be significantly low since our account prohibits treating individuals more harshly than is required to protect society. Just as it is illegitimate to treat carriers of a disease more harshly than is necessary to neutralize the danger they pose, treating criminals more harshly than is required to protect society will be illegitimate as well. A forwarding-looking conception of moral responsibility grounding in future protection and moral formation could justify a suitable fine here, but not more punitive measures. Such small infractions are analogous to common colds. While they do put at risk the health of others, the harm they represent is not significant enough to justify quarantine. Of course, with regard to running a stop sign, we might want to distinguish between first offense and habitual behavior since per incident risk is probably low but aggregates to a high probability of serious harms. Perhaps, then, we could justify increased sanctions over time for repeat offenders, including higher fines and eventually loss of one's drivers license.

On Corrado's second issue, the incapacitation of the dangerous who haven't committed a crime, on our view there are several moral reasons that count against such a policy. As Ferdinand Schoeman (1979) has argued, and Caruso has stressed (2016a), the right to liberty must carry weight in this context, as should the concern for using people merely as means. In addition, the risk posed by a state policy that allows for preventative detention of nonoffenders needs to be taken into serious consideration. In a broad range of societies,

[214] *Derk Pereboom and Gregg D. Caruso*

allowing the state this option stands to result in much more harm than good because misuse would be likely. Schoeman also points out that while the kinds of testing required to determine whether someone is a carrier of a communicable disease may often not be unacceptably invasive, the type of screening necessary for determining whether someone has violent criminal tendencies might well be invasive in respects that raise serious moral issues. Moreover, available psychiatric methods for discerning whether an agent is likely to be a violent criminal are not especially reliable, and, as Stephen Morse points out, detaining someone on the basis of a screening method that frequently yields false positives is seriously morally objectionable (Morse 1999; Nadelhoffer et al. 2012).

However, there is reason to think that impressive neural tests for violent tendencies are being developed (Nadelhoffer et al. 2012). We may in the near future be able to determine with reasonable accuracy on the basis of neural factors whether someone is likely to commit violent crimes. Would our account endorse someone's preventative detention even if he has not yet manifested such violence, on the supposition that the violence would be serious and highly likely in his normal environment and that less invasive measures such as effective monitoring or drug therapy are unavailable? Perhaps it would. But this should not count as a strong objection to our view because virtually everyone would agree that preventative detention of nonoffenders is legitimate under certain possible conditions. Imagine that someone has involuntarily been given a drug that makes it virtually certain that he will brutally murder at least one person during the one-week period he is under its influence. There is no known antidote, and because he is especially strong, mere monitoring would be ineffective. Almost everyone would affirm that it would be at least *prima facie* permissible to preventatively detain him for the week. Now suppose that reliable neural screening reveals that an agent, if left in his normal environment, is virtually certain to engage in rape and murder in the near future. There is no known viable drug therapy, and mere monitoring would be ineffective. Should he be preventatively detained? Here, it is important to understand that the incapacitation account will specify that the circumstances of such detention would not be harsh and that allowing the agent to be reasonably comfortable and to pursue fulfilling projects would be given high priority. But, even here, there are countervailing moral considerations that must be taken into account. In many societies, the danger of misuse posed by allowing the state to preventatively detain even highly dangerous nonoffenders is a grave concern that stands to outweigh the value of the safety provided by such a policy.

A further worry is raised by Lemos, who argues that if criminals were no more deserving of punishment than noncriminals, it would be unjust to expect criminals alone to bear the burden of violent crime prevention, and this tells in favor of lowering the evidentiary standard for criminality,

thereby preventing more crime. The result would involve more people subject to criminal treatment. We offer two responses. The first is that we do not 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 non-threats. Thus the aim of protection is justified by a right with clear bounds, and not by a consequentialist theory on which the bounds are unclear. Second, we noted earlier that we endorse a theory of blame and moral responsibility that does not feature basic desert. Its aims are protection, moral formation, and reconciliation. Moral formation and reconciliation are not relevant in the case of the innocent, but they are for those who have done wrong. To the extent that treatment of criminals also aims at moral formation and reconciliation, these considerations also serve to set the innocent apart from the guilty. It is not clear how these various factors weigh out, but if, taking all of them into consideration, evidentiary standards for criminal liability are lowered, they would not be lowered by much. And this is offset by the fact that criminals would not be treated as harshly as they are in our current system.

On the third issue, cost to society, when a person with cholera is quarantined, she is typically made to experience deprivation she does not deserve. Society benefits by this deprivation. It is a matter of fairness that society do what it can, within reasonable bounds, to make the victim safe for release as quickly as possible, and this will have a cost. If we quarantined cholera victims but were unwilling to provide medical care for them because it would require a modest increase in taxation, then we would be acting unfairly. Similarly, when a dangerous agent, whether or not he has already committed a crime, is preventatively detained, then, supposing that the free will skeptic is right, he is made to experience a deprivation he does not fundamentally deserve and from which society benefits. By analogy with the cholera case, here also it is a matter of fairness for us to do what we can, within reasonable bounds, to rehabilitate him and make him safe for release, and this, too, will have a cost. For a society or state to oppose programs for rehabilitation because it is unwilling to fund them would involve serious unfairness. Corrado suggests that this cost would be prohibitively high. It is hard to see why it would be more costly than the current system, which is massively expensive.

Saul Smilansky (2011) sets out a version of this objection, and Corrado expresses approval. This may explain Corrado's concern about cost:

Hard determinists cannot, however, permit incarceration in institutions of punishment such as those that currently prevail. Instead of punishment, they must opt for funishment. Funishment would resemble punishment in that criminals would be incarcerated apart from lawful society; and institutions of funishment

[216] *Derk Pereboom and Gregg D. Caruso*

would also need to be as secure as current prisons, to prevent criminals from escaping. But here the similarity ends. For institutions of funishment would also need to be as delightful as possible. . . . Since hard determinism holds that no one deserves the hardship of being separated from regular society, this hardship needs to be compensated for. (Smilansky 2011: 355)

Smilansky then argues that such a policy will be extremely expensive: “the cost of funishment will be incomparably higher than that of punishment,” in fact so high that it will be intolerable.

First of all, because the free will skeptic rejects basic desert, no basic desert requirement to compensate those who are preventatively detained will be in effect (cf. Levy 2012; Pereboom 2013, 2014a). The specifics of the skeptic’s reply to Smilansky’s objection depend on which general moral theory she prefers to adopt. Suppose she endorsed an axiological moral theory which includes better consequences as valuable, where morally fundamental rights being honored and not violated count among the good consequences. Neil Levy (2012) points out that a consequentialist of this sort has a good response to Smilansky: “A consequentialist who is a moral responsibility skeptic will naturally hold that no one should be treated any worse than is needed to bring about the best consequences, with all agents’ welfare—including the welfare of criminals—taken into account.” So, first, the preventatively detained would not be treated worse than needed to protect against the danger they pose. In addition, the right to live a fulfilling life is in play and weighs heavily, and we would thus have a significant moral interest in providing those who are preventatively detained with the requisite opportunities and conditions. On the issue of cost, providing these sorts of opportunities may add expense to our system for dealing with criminal behavior, but not the expense required to provide all of those detained with “five-star hotel” accommodations (Smilansky 2011). Furthermore, as Levy (2012) contends, “rejecting the notion that some agents deserve punishment opens the way for us to adopt policies that respond to crime at much lower costs, economically, socially and morally.” Mark Kleiman (2009) proposes and discusses many such less costly nonretributive policies, and he argues persuasively and in detail that adopting them instead of what we in the United States have in place would lead to highly beneficial consequences.

4. CONCLUSION

We have here considered two different routes to free will skepticism and argued that while the first route (the one that denies the causal efficacy of the types of willing required for free will) fails for a number of reasons, the second route based on hard incompatibilism is sound. We then

considered the main concern critics have with this view (i.e., that we cannot live as if it is true) and argued that it can be answered. More specifically, we first argued that there are forward-looking aspects of our practice of holding responsible that don't presuppose basic desert that the skeptic can retain (i.e., those that aim at moral formation, protection, and reconciliation). We further maintained that emotional attitudes that don't implicate basic desert are sufficient for good human relationships. Last, and our primary focus, we argued that a nonretributivist set of practices for treating criminals, which highlight rehabilitation and incapacitation, are in fact workable.

REFERENCES

- Alexander, J. F., and B. V. Parsons. 1982. *Functional family therapy*. Monterrey: Brooks/Cole.
- Bargh, J. A. 1997. The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition* (vol. 10), pp. 1–61. Mahwah, NJ: Erlbaum.
- Bargh, J. A. 2008. Free will is un-natural. In J. Baer, J. C. Kaufman, and R. F. Baumeister (Eds.), *Are we free? Psychology and free will*, pp. 128–154. New York: Oxford University Press.
- Bargh, J. A., and T. L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54 (7): 462–479.
- Bargh, J. A., and M. J. Ferguson. 2000. Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin* 126 (6): 925–945.
- Bentham, J. 1823/1948. *An introduction to the principles of morals and legislation*. New York: Macmillan.
- Berman, M. N. 2008. Punishment and justification. *Ethics* 18: 258–290.
- Baumeister, R. F. 2008. Free will in scientific psychology. *Perspectives of Psychological Science* 3(1): 14–19.
- Bok, H. 1998. *Descartes: His moral philosophy and psychology*. New York: New York University Press.
- Bjornsson, G., and D. Pereboom. 2014. Free will skepticism and bypassing. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (vol. 4). Cambridge, MA: MIT Press.
- Caruso, G. D. 2012. *Free will and consciousness: A determinist account of the illusion of free will*. Lanham, MD: Lexington Books.
- Caruso, G. D. 2014. Précis of Derk Pereboom's *Free will, agency, and meaning in life*. *Science, Religion and Culture* 1(3): 178–201.
- Caruso, G. D. 2015. If consciousness is necessary for moral responsibility, then people are less responsible than we think. *Journal of Consciousness Studies* 22(7–8): 49–60.
- Caruso, G. D. 2016a. Free will skepticism and criminal behavior: A public health-quarantine model. *Southwest Philosophy Review* 32(1): 25–48.
- Caruso, G. D. 2016b. Consciousness, free will, and moral responsibility. (December 21.) Unedited version. <https://ssrn.com/abstract=2888513> Edited version in *Routledge handbook of consciousness*, ed. Rocco J, Gennaro, forthcoming.

[218] *Derk Pereboom and Gregg D. Caruso*

- Caruso, G. D. 2017a. Free will skepticism and the question of creativity: Creativity, desert, and self-creation. *Ergo* 3(23): 23–39.
- Caruso, G. D. 2017b. *Public health and safety: The social determinants of health and criminal behavior*. UK: ResearchLinks Books.
- Caruso, G. D., and S. Morris. 2017. Compatibilism and retributivist desert moral responsibility: On what is of central philosophical and practical importance. *Erkenntnis*. doi: 10.1007/s10670-016-9846-2
- Center for Science and Law. 2012. Deep brain stimulation in rehabilitating criminal psychopaths. <http://www.neulaw.org/blog/1034-class-blog/3972-deep-brain-stimulation-in-rehabilitating-criminal-psychopaths>
- Corrado, M. L. 2016. Two models of criminal justice. <http://ssrn.com/abstract=2757078>
- D'Angelo, E. 1968. *The problem of free will and determinism*. Columbia: University of Missouri Press.
- De Caro, M. 2011. Is emergence refuted by the neurosciences? The case of free will. In A. Corradini and T. O'Connor (Eds.), *Emergence in science and philosophy*, pp. 190–221. London: Routledge.
- Doris, J. M. 2002. *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.
- Evans, J. R. (Ed.). 2006. *Forensic applications of QEEG and neurotherapy*. Informa Healthcare. <https://www.amazon.com/Forensic-Applications-Neurotherapy-James-Evans/dp/0789030799>
- Feinberg, J. 1970. Justice and personal desert. In his *Doing and deserving*. Princeton, NJ: Princeton University Press.
- Fischer, J. M. 2007. Compatibilism. In J. Fischer, R. Kane, D. Pereboom, and M. Vargas (Eds.), *Four views on free will*, pp. 44–84. Hoboken, NJ: Wiley-Blackwell Publishing.
- Fischer, J. M. 2014. Review of *Free will, agency, and meaning in life*, by Derk Pereboom. *Science, Religion and Culture* 1(3): 202–208.
- Gendreau, P., and R. Ross. 1979. Effective correction treatment: Bibliography for cynics. *Crime and Delinquency* 25: 463–489.
- Gkotsi, G-M., and L. Benaroyo. 2012. Neuroscience and the treatment of mentally ill criminal offenders: Some ethical issues. *Journal of Ethics in Mental Health* 6: 1–7.
- Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6): 1464–1480.
- Greteman, B. 2009. Improve mental health with neurofeedback. *OdeWire Magazine*. <http://odewire.com/61556/improve-mental-health-with-neurofeedback.html>
- Haggard, P., and M. Eimer. 1999. On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research* 126(1): 128–133.
- Haji, I. 1998. *Moral accountability*. New York: Oxford University Press.
- Hammond, C. D. 2011. What is neurofeedback: An update. *Journal of Neurotherapy* 15: 305–336.
- Hoerich, M. R. 2011. An analysis of the proposal of deep brain stimulation for the rehabilitation of criminal psychopaths. Presented at Michigan Association of Neurological Surgeons. http://www.destinationmi.com/documents/2011MAN_Spresentation_MarkHoerich.pdf
- Honderich, T. 1988. *A theory of determinism: The mind, neuroscience, and life-hopes*. Oxford: Oxford University Press. Republished in two volumes: *Mind and brain and The consequences of determinism*.
- Husak, D. 2000. Holistic retributivism. *California Law Review* 88: 991–1000.

- Kang, J., and K. Lane. 2010. Seeing through colorblindness: Implicit bias and the law. *UCLA Law Review* 58(2): 465–520.
- King, M., and P. Carruthers. 2012. Moral responsibility and consciousness. *Journal of Moral Philosophy* 9: 200–228.
- Kleiman, M. 2009. *When brute force fails: How to have less crime and less punishment*. Princeton, NJ: Princeton University Press.
- Lemos, J. 2016. Moral concerns about responsibility denial and the quarantine of violent criminals. *Law and Philosophy* 35(5): 461–483.
- Levy, N. 2005. Libet's impossible demand. *Journal of Consciousness Studies* 12(12): 67–76.
- Levy, N. 2011. *Hard luck: How luck undermines free will and moral responsibility*. New York: Oxford University Press.
- Levy, N. 2012. Skepticism and sanctions: The benefits of rejecting moral responsibility. *Law and Philosophy* 31: 477–493.
- Levy, N. 2014. *Consciousness and moral responsibility*. New York: Oxford University Press.
- Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Science* 8:529–566.
- Libet, B. 1999. Do we have free will? *Journal of Consciousness Studies* 6(8–9): 47–57. Reprinted in R. Kane (Ed.), *The Oxford handbook of free will*, pp. 551–564. New York: Oxford University Press, 2002.
- Libet, B., C. A. Gleason, E. W. Wright, and D. K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain* 106: 623–642.
- Losen, D., C. Hodson, M. A. Keith II, K. Morrison, and S. Belway. 2015. Are we closing the school discipline gap? *The Center for Civil Rights Remedies*, February 23.
- Losen, D., T. Martinez, and V. Okelola. 2014. Keeping California's kids in school: Fewer students of color missing school for minor misbehavior. *The Center for Civil Rights Remedies*, June 10.
- Martin, G., and C. L. Johnson. 2005. The boys totem town neurofeedback project: A pilot study of EEG biofeedback with incarcerated juvenile felons. *Journal of Neurotherapy* 9(3): 71–86.
- Mele, A. 2009. *Effective intentions*. New York: Oxford University Press.
- Morse, S. 1999. Neither desert nor disease. *Legal Theory* 5: 265–309.
- Nadelhoffer, T. 2011. The threat of shrinking agency and free will disillusionism. In L. Nadel and W. Sinnott-Armstrong (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet*, pp. 173–188. New York: Oxford University Press.
- Nadelhoffer, T., S. Bibas, S. Grafton, K. A. Kiehl, A. Mansfield, W. Sinnott-Armstrong, and M. Gazzaniga. 2012. Neuroprediction, violence, and the law: Setting the stage. *Neuroethics* 5: 67–99.
- Nahmias, E. 2002. When consciousness matters: A critical review of Daniel Wegner's *The illusion of conscious will*. *Philosophical Psychology* 15(4): 527–541.
- Nahmias, E. 2011. Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford handbook of free will* 2nd ed., pp. 555–576. New York: Oxford University Press.
- Nelkin, D. K. 2014. Free will skepticism and obligation skepticism: Comments on Derk Pereboom's *Free will, agency, and meaning in life*. *Science, Religion and Culture* 1(3): 209–217.
- Nichols, S. 2007. After compatibilism: A naturalistic defense of the reactive attitudes. *Philosophical Perspectives* 21: 405–528.

[220] *Derk Pereboom and Gregg D. Caruso*

- Nisbett, R. E., and T. D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84: 231–259.
- Nosek, B. A., F. L. Smyth, J. J. Hansen, T. Devos, N. M. Linder, K. A. Ranganath, et al. 2007. Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology* 18: 36–88.
- Obhi, S. S., and P. Haggard. 2004. Free will and free won't: Motor activity in the brain precedes our awareness of the intention to move, so how is it that we perceive control? *American Scientist* 92(July–August): 358–365.
- Patterson, G. R., P. Chamberlain, and J. Reid. 1982. A comparative evaluation of a parent training program. *Behavior Therapy* 13: 638–650.
- Pereboom, D. 1995. Determinism *al dente*. *Nous* 29: 21–45.
- Pereboom, D. 2001. *Living without free will*. New York: Cambridge University Press.
- Pereboom, D. 2012. On John Fischer's *Our stories*. *Philosophical Studies* 158: 523–528.
- Pereboom, D. 2013. Free will skepticism and criminal punishment. In Thomas Nadelhoffer (Ed.), *The future of punishment*, pp. 49–78. New York: Oxford University Press.
- Pereboom, D. 2014a. *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Pereboom, D. 2014b. Replies to John Fischer and Dana Nelkin. *Science, Religion, and Culture* 1: 218–225.
- Pockett, S. 2004. Does consciousness cause behavior? *Journal of Consciousness Studies* 11: 23–40.
- Quirk, D. A. 1995. Composite biofeedback conditioning and dangerous offenders: III. *Journal of Neurotherapy* 1(2): 44–54.
- Richmond, E. 2015 December 29. When restorative justice in schools works. *The Atlantic*.
- Roediger III, H. L., M. K. Goode, and F. M. Zaromb. 2008. Free will and the control of action. In J. Baer, J. Kaufman, and R. F. Baumeister (Eds.), *Are we free? Psychology and free will*, pp. 205–225. New York: Oxford University Press.
- Schoeman, F. 1979. On incapacitating the dangerous. *American Philosophical Quarterly* 16: 27–35.
- Scanlon, T. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Sie, M., and A. Wouters. 2010. The BCN challenge to compatibilist free will and personal responsibility. *Neuroethics* 3(2): 121–133.
- Smilansky, S. 2011. Hard determinism and punishment: A practical reductio. *Law and Philosophy* 39: 353–367.
- Smith, P. N., and M. W. Sams. 2005. Neurofeedback with juvenile offenders: A pilot study in the use of QEEG-based and analog-based remedial neurofeedback training. *Journal of Neurotherapy* 9(3): 87–99.
- Sommers, T. 2007. The objective attitudes. *Philosophical Quarterly* 57: 321–341.
- Soon, C. S., M. Brass, H.-J. Heinze, and J.-D. Haynes. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* 11(5): 543–545.
- Strawson, G. 1994. The impossibility of moral responsibility. *Philosophical Studies* 75(1): 5–24.
- Strawson, P. F. 1962. Freedom and resentment. *Proceedings of the British Academy* 48:1–25. Reprinted in D. Pereboom (Ed.), *Free will*, pp. 119–142. Indianapolis, IN: Hackett.
- Uhlmann, E. L., and G. L. Cohen. 2005. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science* 16: 474–480.

- US Department of Education Office for Civil Rights. 2014 March 1. Civil rights data collection: Data snapshot: School Discipline. Issue Brief No. 1. <http://ocrdata.ed.gov/Downloads/CRDC-School-Discipline-Snapshot.pdf>
- Waller, B. 1990. *Freedom without responsibility*. Philadelphia: Temple University Press.
- Waller, B. 2011. *Against moral responsibility*. Cambridge, MA: MIT Press.
- Walters, G. D. 1992. *Foundations of criminal science*. New York: Praeger.
- Wegner, D. M. 2002. *The illusion of conscious will*. Cambridge, MA: Bradford Books, MIT Press.
- Wilson, T. 2002. *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Yochelson, S., and S. Samenow. 1976. *The criminal personality: A profile for change*. New York: Aronson.