

## Article

## Theme: AUTHOR MEETS CRITICS

Précis of Derk Pereboom's *Free Will, Agency, and Meaning in Life*

Gregg D. Caruso

Corning Community College (SUNY)

Email: gcaruso@corning-cc.edu

Derk Pereboom's *Free Will, Agency, and Meaning in Life* (2014) provides the most lively and comprehensive defense of free will skepticism in the literature. It contains a reworked and expanded version of the view he first developed in *Living without Free Will* (2001). Important objections to the early book are answered, some slight modifications are introduced, and the overall account is significantly embellished—for example, Pereboom proposes a new account of rational deliberation consistent with the belief that one's actions are causally determined (ch.5) and develops a forward-looking theory of moral responsibility consistent with free will skepticism (ch.6). A significant contribution to the field, *Free Will, Agency, and Meaning in Life* is destined to become a classic and is essential reading for anyone interested in free will and moral responsibility.

The goal of the book is to advocate for *free will skepticism*: the view 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*—the sense that would make us *truly deserving* of blame or praise. Pereboom follows tradition here in defining “free will” in terms of the control in action required for a core sense of moral responsibility. “This sense of moral responsibility, the one at issue in the free will debate, is set apart by the notion of *basic desert* (Feinberg 1970; Pereboom 2001, 2007a; G. Strawson 1994; Fischer 2007: 82; Clarke 2005; Scanlon 2013). 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” (2). Such desert is ba-

sic 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” (2).

Pereboom's argument against this sort of free will is known as *hard incompatibilism*, which amounts to a rejection of both compatibilism and libertarianism. Hard incompatibilism 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. In addition to hard incompatibilism, Pereboom also defends *optimistic skepticism*—which maintains that life without free will and desert-based moral responsibility would not be as devastating to our conceptions of agency, morality, and meaning in life as some suggest, and in certain respects it may even be beneficial. “In particular, this conception is wholly compatible with rational deliberation, with practically viable notions of morality and moral responsibility, with a workable system of dealing with criminal behavior, and with a secure sense of meaning in life” (4).

Pereboom's overall argument can be sketched as follows: Against the view that free will is compatible with the causal determination of our actions by natural factors beyond our control, Pereboom argues that there is no relevant difference between this prospect and our actions being causally determined by manipulators. Against event causal libertarianism, he advances the disappearing agent objection, according to which on this view the agent cannot settle whether a decision

occurs, and hence cannot have the control required for moral responsibility. Against a non-causal libertarianism, he likewise argues that it has no plausible proposal of the control in action required for responsibility. He goes on to argue that while agent-causal libertarianism may supply this sort of control, it cannot be reconciled with our best physical theories. Since this exhausts the options for views on which we have the sort of free will at issue, he concludes that free will skepticism is the only remaining position. Finally, Pereboom defends the optimistic view that conceiving of life without free will would not be devastating to our conceptions of agency, morality, and meaning in life, and may even be beneficial.

In this précis I will provide a detailed summary of the arguments that lead Pereboom to these conclusions—including his disappearing agent objection (ch.2), his argument against agent-causal libertarianism (ch.3), his four-case argument (ch.4), and his defense of optimistic skepticism (ch.5-8). My goal is simply to present Pereboom's arguments in a clear and concise manner. Providing a comprehensive summary of *Free Will, Agency, and Meaning in Life*, without any additional commentary or critical comments, will allow readers to evaluate Pereboom's arguments for themselves. It will also allow the articles to follow in this book symposium—those by John Martin Fisher, Dana Nelkin, and Derk Pereboom—to provide their comments without having to summarize the arguments in full. Readers can then refer back to this précis for additional clarification or details on any particular argument.

## Chapter 1: Defending a Source View

Pereboom begins by arguing for a source view of moral responsibility by way of a Frankfurt-style case he first developed in *Living without Free Will*. Before laying out that case, however, it is important to first draw some distinctions and define some terminology. *Source* views contend that an agent's being morally responsible for an action (in the basic desert sense) is, or would be, explained primarily by the action's having a causal history in which she is the source of her action in a specific way. Such views are contrasted with *leeway* accounts, which instead maintain that an agent's moral responsibility for an action is explained by the availability to her of alternative possibilities, for example by the ability to refrain from doing what she has actually done.

Additionally, a source theorist might be a *compatibilist* (Fischer 1994; Fischer and Ravizza 1998; McKenna 2005; Sartorio 2011, 2013) or an *incompatibilist* (Stump 1990, 1996; Zagzebski 1991, 2000; Hunt 2000, 2005; Shabo 2010); just as a leeway theorist may be a compatibilist (Moore 1912; Ayer 1954; Vihvelin 2013) or incompatibilist (van Inwagen 1983; Kane 1996). *Incompatibilists* maintain that the causal determination of our actions by factors beyond our control is incompatible with the sort of free will required for 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*). *Compatibilists*, on the other hand, maintain that determinism is in fact compatible with the ability to do otherwise (*leeway compatibilism*) or with the agent's being the source of her action in the required way (*source compatibilism*).

Pereboom is a *source incompatibilist*. He rejects the sort of incompatibilism for which the availability of alternative possibilities is crucial to explaining moral responsibility in the basic desert sense, and accepts instead an incompatibilism that ascribes the more significant role to an action's causal history, to the actual causal sequence that produces it (Fischer 1982, 1994, 2007; Fischer and Ravizza 1998; Sartorio 2011). As Pereboom describes it:

My view is thus a *source* as opposed to a *leeway* incompatibilism. Agent-causal libertarianism is commonly conceived as an incompatibilist position in which an agent can be the source of her action in the way required for moral responsibility in the sense at issue, and as a result proponents of this view are typically source incompatibilists. But one might also be a source incompatibilist and seriously doubt that we have the sort of free will required for this sort of moral responsibility, and this is the position I advocate. (4)

Pereboom's free will skepticism is therefore a result of his source incompatibilism combined with his rejection of libertarianism. Pereboom lays out his arguments against libertarianism in chapters 2 and 3, and presents his manipulation argument against compatibilism in chapter 4. The goal of chapter 1, however, is simply to defend the source view against recent ob-

jections.

In recent decades, and largely due to Frankfurt-style examples (Frankfurt 1969) against the Principle of Alternative Possibilities (PAP), it has become common to claim that an agent's moral responsibility for an action is not explained by the availability to her of alternative possibilities but rather by the agent's being the actual source of her action in a specific way (9). Some philosophers, however, have resisted this trend and have criticized Frankfurt-style cases by presenting the dilemma defense of the leeway view (Kane 1985; Widerker 1995; Ginet 1996). In this chapter, Pereboom defends his Frankfurt-style case, known as the *Tax Evasion Case*, against such objections. In addition, he argues that "this sort of case yields an effective response to a different sort of objection which was originally raised by Ginet and recently developed by Christopher Franklin (2011) and David Palmer (2011), which focuses on the precise timing of what it is that the agent is responsible for" (5).

In typical Frankfurt cases, an agent considers performing some action, but a neuroscientist is concerned that she will not come through. So if she were to manifest an indication that she will not or might not perform the action, the neuroscientist would intervene. But as things actually go, the neuroscientist remains idle, since the agent performs the action on her own. "The idea is that even though the agent could not have avoided the action she performs, she is still intuitively morally responsible for this action" (9). While Frankfurt cases have been extremely influential in moving people away from leeway and toward source accounts of moral responsibility, Robert Kane (1985), David Widerker (1995), and Carle Ginet (1996) have developed a significant objection against them.

According to the Kane/Widerker/Ginet objection, Frankfurt cases face a dilemma: either the proponents of Frankfurt-style examples are presupposing the truth of causal determinism (in which case they are begging the question against the incompatibilist) or they are presupposing the truth of indeterminism (in which case they are wrong in assuming the agent could not have chosen otherwise). As Pereboom summarizes the objection: "In Frankfurt examples the actual situation will feature a prior sign that signals the fact that intervention is not required. If in a proposed case the prior sign causally determined the action, or if it were associated with some factor that did, the in-

tervener's (or his device's) predictive ability could be explained. However, then the libertarian would not and could not be expected to have the intuition that the agent is morally responsible. But if the relationship between the prior sign and the action was not causally deterministic in such ways, then it's open that the agent could have done otherwise despite the occurrence of the prior sign" (2014, 14). Critics therefore maintain that, either way, an alternative-possibilities condition on moral responsibility emerges unscathed.

To understand Pereboom's response to this objection, it's important to first understand his criterion for *robustness*. "The core intuition that underlies the proposal to explain moral responsibility by access to alternative possibilities is that to be blameworthy for an action, an agent must have been able to do something that would have resulted in her being 'off the hook' (Pereboom 2001: 1)" (10-11). Put differently, "to be blameworthy for an action an agent must have an *exempting* alternative possibility, one that, should he avail himself of it, would exempt him from blame (Moya 2006: 67)" (11). In earlier work, Pereboom spelled this out in terms of robustness—that for an alternative possibility to be robust, it must be one that, had the agent taken it instead, would have precluded him from the particular responsibility he actually has for his action. Robustness also has an epistemic component: if an agent were blameworthy because he had an alternative possibility available to him, it must be that he in some sense understood that or how it was available to him (11).

Trying to capture these two components, Pereboom originally stated the robustness criterion as follows—this was meant as a substantial necessary condition for what it is for an alternative possibility to be robust:

**Robustness (A):** For an agent to have a robust alternative to her action A, that is, an alternative relevant per se to explaining why she is morally responsible for A, she must have understood that instead she could have voluntarily done something as a result of which she would have been precluded from the moral responsibility she actually has for A. (11)

But due to objections raised by Jonathan Vance, Dana Nelkin, and others, Pereboom acknowledges in chapter 1 that a more precise criterion for robustness is needed. After considering four main concerns for the

Robustness (A) formulation, he arrives at the following new criterion:

**Robustness (B):** For an agent to have a robust alternative to her immoral action A, that is, an alternative relevant per se to explaining why she is blameworthy for performing A, it must be that

1. She instead could have voluntarily acted or refrained from acting as a result of which she would be blameless, and
2. For at least one such exempting acting or refraining, she was cognitively sensitive to the fact that she could so voluntarily act or refrain, and to the fact that if she voluntarily so acted or refrained she would then be, or would likely be, blameless. (13)

We can now characterize leeway views, whether compatibilist or incompatibilist, “as those that affirm that for an agent to be blameworthy for an action, she must have available to her a robust alternative possibility, that is, one that satisfies Robustness (B)” (13). Additionally, leeway incompatibilists hold that “the reason causal determinism precludes moral responsibility is that it rules out alternative possibilities altogether, but most importantly, those of the robust sort” (14). Source views, by contrast, deny this.

Pereboom, being a source theorist, now introduces his *Tax Evasion Cases*—a modified Frankfurt-style example designed to avoid the dilemma reply.

**Tax Evasion (2):** Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons, of which he is aware. He could not, for example, fail to make this choice for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily.

However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain which, were it to sense the requisite level of attentiveness, would electronically simulate the right neural centers so as to inevitably result in his making this choice. As it happens, Joe does not attain this level of attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle (Pereboom 2000, 2001, 2003, 2009b). (15)

Pereboom maintains that in this situation, Joe is intuitively blameworthy for choosing to evade taxes despite the fact that he did not have a robust alternative (15). And because the case includes a number of features that distinguish it from earlier Frankfurt cases, it is claimed to avoid the dilemma outlined above. The unique features are these: “the cue for intervention is a necessary condition for the agent's availing herself of any robust alternative possibility (without the intervener's device in place), while the cue for intervention itself is not a robust alternative possibility, and the absence at any specific time of the cue for intervention in no sense causally determines the action the agent actually performs” (15).

In *Tax Evasion (2)*, it's not merely the presence of a necessary condition for doing otherwise that's important. Rather, “this case features a necessary condition for doing otherwise *the absence of which at any specific time will not causally determine the agent to perform the action*” (15). According to Pereboom, this allows it to be the case that “at no specific time is the agent causally determined to perform the action, which facilitates satisfaction of a key libertarian condition on moral responsibility” (15). Here's how Pereboom describes it:

In *Tax Evasion (2)*, the necessary condition for Joe's not deciding to evade taxes, i.e., his having the requisite level of attentiveness to the moral reasons, is the right sort, since its absence at any specific time does not causally determine his deciding to evade taxes. For at any particu-

lar time at which the level of attentiveness is absent, Joe could still make it occur at a later time, and thus he is not causally determined to decide to evade taxes by its absence at the previous time...(15)

It's also important to note that the case does feature alternative possibilities that are available to the agent—Joe's achieving higher levels of attentiveness to moral reasons—but Pereboom argues that these alternatives are not robust.

At this point one might object that given that the intervener's device is in place, by voluntarily achieving the specified higher level of attentiveness Joe would have voluntarily done something as a result of which he would have been precluded from the blameworthiness he actually incurs. Had he voluntarily achieved the requisite level of attentiveness, the intervention would have taken place, whereupon he would not have been blameworthy for deciding to evade taxes. But this alternative possibility is not robust. Joe is not even minimally cognitively sensitive to the fact that by voluntarily achieving the requisite level of attentiveness he would not be (or would likely not be) blameworthy. Moreover, he has no reason whatsoever to believe this. True, were he voluntarily to attain this attentiveness, the intervention would take place, and he would then not have been blameworthy. Still, Joe has no inkling, and no evidence, that the intervention would then take place and that as a result he would not be blameworthy. (16)

According to Pereboom, Joe believes that achieving this level of attentiveness is compatible with his freely deciding to evade taxes anyway, and he has no reason to suspect otherwise. "We can even specify that he believes that if he did achieve this level of attentiveness, he would still be very likely to decide to evade taxes" (16). Intuitively, then, Joe is blameworthy for actually so deciding.

The remainder of ch.1 is dedicated to addressing objections to the Tax Evasion (2) Case. Pereboom responds to objections by Widerker (2006), Kane (2007), Ginet (2002), and others, and argues that each of them can be adequately dealt with. He concludes that there is good reason to reject the leeway

view and adopt a source view, which maintains that for an agent to be morally responsible for an action she must be the source of her action in an appropriate way. More specifically, he embraces source incompatibilism, which contends that this requires the action's actual causal history be indeterministic—i.e., "the action cannot be causally determined by factors beyond the agent's control, such as the remote past together with the laws of nature" (28).

## Chapter 2: Problems for Event-Causal and Non-Causal Libertarianism

After having defended the source view in ch.1, Pereboom turns his attention in ch.2 to arguing against two leading varieties of libertarianism. In particular, Pereboom develops his "disappearing agent" objection against *event-causal libertarianism*, which argues that this position cannot accommodate the requirement that in an indeterministic context the agent or something about the agent settles whether the decision in question occurs, and for this reason lacks the control in action required for moral responsibility in the basic desert sense (5). In *Living without Free Will*, Pereboom directed this objection toward Robert Kane's (1996) famed account, but here he focuses instead on Mark Balaguer's (2009) version of event-causal libertarianism. Pereboom also takes the opportunity in this chapter to argue against *non-causal libertarianism*, objecting that advocates of this position "use *prima facie* causal language to express the purportedly non-causal relation, and that either causation is being invoked, or if it is not, the control required for moral responsibility is absent" (40).

Pereboom begins by differentiating three major versions of libertarianism: event-causal, non-causal, and agent-causal types (30-31). Libertarian views, in general, maintain that human beings have the capacity to freely will actions, and crucial to an agent's being freely willed is that it not be causally determined by factors beyond the agent's control. *Event-causal libertarianism*, in particular, maintains that actions (conceived as agent-involving events) are caused solely by prior events, such as an agent's having a desire or a belief at a time, and some type of indeterminacy in the production of action by appropriate events is held to be necessary for the sort of free will under discussion (Kane 1996; Ekstrom 2000; Balaguer 2009). *Non-causal libertarianism*, on the other hand, maintains that no cause is needed for human decisions. Contemporary

non-causal theorists include Carl Ginet (1990, 1996, 2007), Hugh McCann (1998), and Stewart Goetz (2008). Typically, non-causal theorists also maintain that additional conditions must be met for an action to be free. *Agent-causal libertarianism*, the most metaphysically demanding of the three views, appeals to agents who as substances have the power to cause decisions without being causally determined to do so (Kant 1781/1787/1987; Reid 1788/1983; Taylor 1966, 1974; Chisholm 1964, 1976; O'Connor 2000, 2009; Clarke 1993, 2003; Griffith 2010).

Pereboom notes that “event-causal libertarianism is typically regarded as the most attractive of these views, all else being equal.” And this is because the “idea of an uncaused event and the notion of a substance-cause are regarded with suspicion” (2014, 31). Yet despite its attractiveness, event-causal libertarianism faces a devastating objection known as the *disappearing agent objection*. The objection runs as follows:

*The disappearing agent objection:* Consider a decision that occurs in a context in which the agent’s moral motivations favor that decision, and her prudential motivations favor her refraining from making it, and the strengths of these motivations are in equipoise. On an event-causal libertarian picture, the relevant causal conditions antecedent to the decision, i.e., the occurrence of certain agent-involving events, do not settle whether the decision will occur, but only render the occurrence of the decision about 50% probable. In fact, because no occurrence of antecedent events settles whether the decision will occur, and only antecedent events are causally relevant, *nothing* settles whether the decision will occur. Thus it can’t be that the agent or anything about the agent settles whether the decision will occur, and she therefore will lack the control required for basic desert moral responsibility for it (Pereboom 2001, 2004, 2007a). (32)

The concern raised by Pereboom’s disappearing agent objection is that “because event-causal libertarian agents will not have the power to settle whether the decision will occur, they cannot have the role in action that this sort of moral responsibility demands” (32).

Consider, for instance, Balaguer’s event-causal libertarian account. As Balaguer describes it, his view

“does not involve any sort of irreducible agent causation, but it does hold that undetermined L-free decisions are (ordinarily) causally influenced by—indeed, probabilistically caused by—agent-involving events, most notably, events having to do with the agent having certain reasons and intentions” (2009, 67). For Balaguer, the paradigm case of a free choice is one in which the agent’s reasons and motivation are in equipoise. He calls this a *torn decision*: a decision in which the agent (a) has reasons for two or more options and feels torn as to which set of reasons is stronger, that is, has no conscious belief as to which option is best, given her reasons; and that (b) decides without resolving the conflict—that is, the person has the experience of “just choosing” (Balaguer 2009, 71).

This definition of a torn decision sets up Balaguer’s conception of libertarian freedom—*L-freedom*:

If an ordinary human torn decision is wholly undetermined, then it is *L-free*—that is, it is not just undetermined but also appropriately non-random, and the indeterminacy increases or procures the appropriate non-randomness. (2009, 78)

Balaguer, in turn, characterizes appropriate non-randomness in terms of authorship and control: “in order for a decision to be L-free, it has to be authored and controlled by the agent in question; that is, it has to be her decision, and she has to control which option is chosen” (2009, 83). Bringing all the pieces together, Balaguer defines the notion of *torn-decision indeterminism*, or TDW-indeterminism, as follows:

*TDW-indeterminism:* Some of our torn decisions are wholly undetermined at the moment of choice, that is, the moment-of-choice probabilities of the various reasons-based tied-for-best option match the reasons-based probabilities, so that these moments-of-choice probabilities are all roughly even, given the complete state of the world and the laws of nature, and the choice occurs without any further input, that is, without anything else being significantly causally relevant to which option is chosen. (2009, 34)

In the case of a torn TDW-indeterminist decision, then, there is no mismatch between the underlying probabilities for the various options at the time of the

decision and the probabilities for those options at that time based on the agent's consciously available reasons (34).

While this account does a nice job articulating a plausible set of epistemic and metaphysical conditions necessary for event-causal libertarianism, Pereboom objects that it does not adequately deal with the disappearing agent objection since it fails to secure the control necessary for basic desert moral responsibility. "Intuitively, this sort of control requires the agent or at least something about the agent to settle whether the decision in question will occur, and the event-causal libertarian view does not allow for this in the case of torn decisions" (34). Pereboom's objection therefore maintains that "on the event-causal libertarian view, torn decisions cannot, in Balaguer's terminology, be appropriately non-random, and the indeterminacy in question cannot increase or procure the appropriate non-randomness, because the theory does not allow either the agent or anything about the agent to settle whether such a decision will occur, and so the requisite control in action is precluded." Moreover, adds Pereboom, "authorship is missing given that such control is required for authorship" (35).

The remainder of ch.2 is dedicated to (a) defending the disappearing agent objection against possible replies, (b) extending the basic concern to non-causal views, (c) considering a possible mixed theory—one that includes event-causal and non-causal components, and (d) arguing against a phenomenological defense of libertarianism.

Against non-causal views, Pereboom argues that the control required for desert-based moral responsibility is once again absent. He also notes how advocates of such views use *prima facie* causal language to express the purportedly non-causal relation, compromising their non-causalism (see pages 40-41). Mixed theories, which incorporate components of both event-causal and non-causal accounts, fair no better according to Pereboom—and this is because they either smuggle in causal notions (in ways they deny) through a "making happen" or "difference-making" relation, or they become incoherent (43-45). Finally, Pereboom considers a phenomenological defense of event-causal libertarianism, which cites phenomenology in support of the claim that event-causal agents have the requisite kind of control in deciding (see, e.g., Balaguer 2009, 89-91; Kane 1996). Pereboom acknowledges that "when we

make decisions we often have the phenomenology of control in acting, and perhaps even of the sort of control or free will that's required for moral responsibility in the basic desert sense." Yet in the free will debate, "serious questions are raised about whether a theory of agency that is causally deterministic or one that endorses pure event-causal indeterminism can accommodate the veridicality of such phenomenology of control." Pereboom therefore concludes that, "In the absence of a satisfying response to this concern, citing the phenomenology in support of control should not [be] counted as effective" (47).

The conclusion of ch.2 is that event-causal and non-causal libertarian accounts both fail to secure the control in action necessary for basic desert moral responsibility. Chapter 3, to which I will now turn, provides Pereboom's argument against agent-causal libertarianism. The conclusion reached at the end of ch.3 is that either a compatibilist view of free will is viable, or else free will skepticism is true. But chapter 4 is devoted to setting out a manipulation argument against any form of compatibilism. Hence, chapters 2-4 collectively represent Pereboom's argument for free will skepticism.

### Chapter 3: The Prospects for Agent-Causal Libertarianism

Agent-causal libertarianism attempts to secure the control missing from event-causal accounts by reintroducing the agent as a cause, not merely as involved in events, but rather fundamentally as a substance. Agent-causal theorists posit an agent who possesses a causal power, fundamentally as a substance, to cause a decision—or as O'Connor (2009) specifies, "the coming to be of a state of intention to carry out some action"—without being causally determined to do so, and thereby to settle, with the requisite control, whether this state of intention will occur (51). Pereboom maintains that agent-causal accounts are "substantive" and can withstand the kind of concerns outlined in the previous chapter, including the *luck objection* (cf. Mele 1999, 2006; Haji 2004). Agent-causal accounts, Pereboom maintains, "deliver what the luck objection indicates is lacking in the event-causal libertarian's account" (55)—i.e., involvement of the agent in the making of her decision.

Yet despite the advantage agent-causal accounts have over event-causal accounts in specifying the control

needed for moral responsibility, they too are ultimately unsatisfying (but for a different reason). Pereboom's criticism of agent-causal libertarianism is rather straightforward: "our best empirical theories yield strong reasons to doubt that we are in fact agent causes of the sort that this theory specifies" (50). These reasons concern whether agent-causal libertarianism can be reconciled with what we would expect given our best physical theories. According to agent-causation, when an agent-cause makes a free decision, she causes it without being causally determined to do so. "But she at some point would affect the physical world distinct from the agent-cause" (66). Given that on our best physical theories the physical world is law-governed, the challenge is to reconcile agent-causation with the claim that the laws of physics govern the physical components of human action.

There are two distinct options for an agent-causal libertarianism to attempt this reconciliation, but neither seems satisfying:

Suppose first that the physical laws are deterministic, as they are, for example on Kant's view. In his agent-causal picture, when an agent makes a free decision, she causes the decision without being causally determined to do so. On the path to action that results from this undetermined decision, alterations in the physical world, for example in her brain or some other part of her body, are produced. But it would seem that we would at this point encourage divergences from the deterministic laws. For the alterations in the physical world that result from the undetermined decisions would themselves not be causally determined, and they would thus not be governed by deterministic law. One might object that it is possible that the physical alterations that result from free decisions just happens to dovetail with what could in principle be predicted on the basis of the deterministic laws, so nothing actually occurs that diverges from these laws. But this proposal would, at least *prima facie*, involve coincidences too wild to be credible. For this reason, it seems that agent-causal libertarianism is not reconcilable with the physical world's being governed by deterministic laws. (66)

determinacy could help reconcile this problem (see, e.g., Clarke 1993, 2003; O'Connor 2000, 2009)? As Pereboom notes, "On one interpretation of quantum mechanics, the physical world is not in fact deterministic, but is rather governed by laws that are fundamentally merely probabilistic or statistical" (66). Let us suppose, then, that significant quantum indeterminacy percolates up to neural indeterminacy at the level of decision or intention-formation.

Still, it appears that wild coincidences would also arise on this suggestion. Consider the class of possible human actions each of which has a physical component whose antecedent probability of occurring is approximately 0.32... Are free choices on the agent-causal libertarian model compatible with what the statistical laws lead us to expect about them? If agent-caused free action were compatible with what according to the statistical law is overwhelmingly likely, then for a large enough number of instances the possible actions in our class would have to be freely chosen close to 32 percent of the time. Then, for a large enough number of instances, the possible actions whose physical components have an antecedent probability of 0.32 would almost certainly be freely chosen close to 32 percent of the time. But if the occurrence of these physical components were settled by the choices of agent-causes, then their actually being chosen close to 32 percent of the time would amount to a coincidence no less wild than the coincident of possible actions whose physical components have an antecedent probability of about 0.99 being chosen, over a large enough number of instances, close to 99 percent of time. The proposal that agent-caused free choices do not diverge from what the statistical laws predict for the physical components of our actions would run so sharply counter to what we would expect as to make it incredible (Pereboom 1995, 2001). (67)

Given that both options lead to wild coincidences, coincidences not to be expected and without an explanation, Pereboom concludes that agent-causal libertarianism is unsatisfying.

Hence, the conclusion of chapters 2 and 3 combined is that "while we might hope that we are free agents

What if it was suggested instead that quantum in-



of the libertarian kind, we shouldn't regard this option as credible" (70). This leaves compatibilism as the remaining view that stands to vindicate the belief that we have the sort of free will required for moral responsibility in the basic desert sense.

## Chapter 4: A Manipulation Argument against Compatibilism

Compatibilism maintains that the sort of free will required for desert-based moral responsibility is compatible with an agent's being causally determined to act by factors beyond her control. Such a view has been so attractive that "a large proportion of contemporary philosophers classify themselves as compatibilists" (71). Pereboom identifies two prominent routes to compatibilism. The first and more common route aims to differentiate causal circumstances of actions that exclude moral responsibility from those that do not. "The core idea is that moral responsibility requires some type of causal integration between the agent's psychology and her action, while it does not demand the absence of causal determination" (71). The second route, the one advocated by P. F. Strawson (1962), "specifies that despite what incompatibilists suppose, the truth of determinism is irrelevant to whether we have the sort of free will required for moral responsibility" (72).

In the past, it was common for compatibilists to defend a conditional account of the ability to do otherwise (Hume 1748/2000; Hobbes 1654; Moore 1912; Ayer 1954). According to such accounts, to say that I could have acted otherwise is to claim that I would have acted otherwise *if I had so chosen*. But if Pereboom's defense of the source view in ch.1 succeeds and a successful Frankfurt-style example is available, then "alternative possibilities will not play the prominent role...these conditional accounts suggest" (72). According to Pereboom: "If the availability of alternative possibilities is not crucial to the sort of free will required for moral responsibility, the Consequence Argument, which aims to show that if determinism is true, no agent can ever do otherwise, can be sidestepped..." (73). It is for this reason that Pereboom thinks the best anti-compatibilist strategy involves an argument from manipulation. The core idea of manipulation arguments is that "an action's being produced by a deterministic process that traces back to factors beyond the agent's control, even when she satisfies all the causal conditions on moral responsibility speci-

fied by the contending compatibilist theories, presents in principle no less of a threat to moral responsibility than does deterministic manipulation by other agents" (73).

Pereboom's manipulation argument is known as the *four-case argument*. The argument sets out three examples of actions that involve manipulation, "the first of which features the most radical sort of manipulation consistent with the proposed compatibilist conditions and with intuitive conditions on agency, each progressively more like the fourth, which the compatibilist might envision to be ordinary and realistic, in which the action is causally determined in a natural way" (74-75). The challenge is for the compatibilist to point out a relevant and principled difference between any two adjacent cases that would show why the agent might be morally responsible in the latter example but not in the earlier one. Pereboom maintains that "this can't be done, and that the agent's non-responsibility therefore generalizes from the first of the manipulation examples to the ordinary case" (75). Furthermore, since the first three cases set out examples of actions that involve manipulation, and in which the prominent compatibilist causal conditions on moral responsibility are satisfied, they also indicate that these conditions are inadequate—"that is, they are not, together with some other uncontroversial necessary conditions for moral responsibility, sufficient for it" (74).

Here's the set-up (75-76): In each of the four cases Professor Plum decides to murder White for the sake of some personal advantage, and succeeds in doing so. The action under consideration is his decision to kill White—"a basic mental action" (75). Pereboom specifies that in each of the four cases, Plum's decision to kill White satisfies the relevant compatibilist-friendly conditions for acting freely. For example, it meets certain compatibilist conditions proposed by Hume: "it is not out of character, since for Plum it is generally true that selfish reasons weigh heavily—too heavily when considered from the moral point of view—while in addition the desire that motivates him to act is nevertheless not irresistible for him, and in this sense he is not constrained to act" (75). The action also meets the compatibilist condition proposed by Harry Frankfurt (1971): "Plum's effective desire (i.e., his will) to murder White conforms appropriately to his second-order desires for which effective desire he will have. That is, he wills to murder her, and wants to will to do so" (75). In addition, the action satisfies the

reasons-responsiveness condition advocated by John Fischer and Mark Ravizza (1998): “Plum’s desire can be modified by, and some of them arise from, the rational consideration of his reasons, and if he believed that the bad consequences for himself that would result from his killing White would be more serve than he actually expects them to be, he would not have decided to kill her” (75). Finally, this action also satisfies the related condition advanced by Jay Wallace (1994): “Plum has the general ability to grasp, apply, and regulate his actions by moral reasons. For instance, when egoistic reasons that count against acting morally are weak, he will typically act for moral reasons instead” (75).

The question, then, is that supposing Plum is causally determined by factors beyond his control to decide as he does, is it plausible that he is morally responsible for his decision? The following four cases exhibit varying ways in which Plum’s decision to kill White might be causally determined by factors beyond his control. Case 1 involves manipulation by a team of neuroscientists.

**Case 1:** A team of neuroscientists has the ability to manipulate Plum’s neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reason process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum’s effective first-order desire to kill White conforms to his second-order desire. In addition, his process of deliberation from which the decision results is reason-responsive; in particular, this type of process would have resulted in Plum’s refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum

is also not constrained to act as he does, for he does not act because of an irresistible desire—the neuroscientists do not induce a desire of this sort. (76-77)

It’s intuitive, argues Pereboom, that Plum is not morally responsible in Case 1 despite the fact that his actions satisfy each of the compatibilist conditions. And Plum’s lack of moral responsibility is explained (at least on one candidate explanation) by the fact that his decision is causally determined by the neuroscientists’ intervention, which is beyond his control, together with the fact that he would not have decided to kill White had this intervention not occurred (77).

The next case is more like the ordinary situation than Case 1 but it still involves manipulation.

**Case 2:** Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at times strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different). (77)

Pereboom maintains that here again Plum is not morally responsible for his decision since “it would seem unprincipled to claim that here, by contract with Case 1, Plum is morally responsible because the length of time between the programming and his decision is now great enough” (77-78). It’s irrelevant, argues Pereboom, whether the programming occurs a few seconds before or forty years prior to the action. What is of relevance is that in both cases Plum’s decision to kill White is causally determined by factors beyond his control: “Causal determination by what the neuroscientists do, which is beyond his control, plausibly

explains Plum's not being morally responsible in the first case, and it's intuitive that he is not morally responsible in the second case for the same reason" (78). Pereboom further notes that if Plum lacks moral responsibility in Case 2, this once again indicates that the compatibilist conditions specified, either individually or in conjunction, are not sufficient for moral responsibility.

Case 3 is more similar yet to our ordinary situation. It supposes that Plum was brought up in an environment in which self-interest and violence are more strongly encouraged than they are in ours, even though morality also has a part.

**Case 3:** Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning process so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to prevent or alter these practices. Due to the aspect of his character produced by this training, in his present circumstances he is causally determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White. While Plum does have the general ability to regulate his behavior by moral reasons, in virtue of this aspect of his character and his circumstances he is causally determined to make his immoral decision, although he does not decide as he does due to an irresistible desire. The neural realization of his deliberative reasoning process and of the decision is just as it is in Cases 1 and 2. (78)

The challenge for the compatibilist is to explain how Plum could be morally responsible in Case 3 but fail to be morally responsible in Case 2. To successfully do this he must identify a feature of these scenarios that would explain the difference. But this is impossible, argues Pereboom, since "there is no such feature" (78). Hence, Plum's exemption from responsibility in Cases 1 and 2 generalizes to the near-to-normal Case 3.

Case 4 deals with ordinary determinism. It differs from the previous cases in that other agents do not

bring about Plum's decision, but in all other relevant respects it is similar.

**Case 4:** Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1-3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum's reasoning process and decision is exactly as it is in Cases 1-3; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill. (79)

Pereboom argues that Plum's exemption from moral responsibility generalizes to this ordinary case because "there are no differences between Case 3 and Case 4 that would justify the claim that Plum is not responsible in Case 3 but is in Case 4" (79). The fact that in Case 4 other agents do not bring about the causal determination of Plum's decision is not a relevant difference according to Pereboom. Imagine, for example, further cases that are exactly the same as Case 1 or Case 2, "except that states at issue are instead produced by a spontaneously generated machine—a machine with no intelligent designer (Pereboom 2001: 115) or a force field (Mele 2005)" (79). Since Plum would lack moral responsibility here as well, Pereboom concludes "causal determination by other agents was not essential to what was driving the intuition of non-responsibility in the earlier cases" (79).

For Pereboom, the best explanation for Plum's non-responsibility in all four cases is that, in each, he is causally determined by factors beyond his control to decide as he does. And since Plum is not morally responsible in Case 1, and there are no differences between Cases 1 and 2, 2 and 3, and 3 and 4 that can explain in a principled way why he would not be responsible in the former of each pair but would be in the latter, "[w]e are thus driven to the conclusion that he is not responsible in Case 4" (79). This, then, is Pereboom's four-case argument and it represents his reason for rejecting compatibilism.

The remainder of ch.4 is dedicated to addressing objections to the manipulation argument, especially those of Alfred Mele, John Martin Fischer, Michael McKenna, and Daniel Haas. Roughly speaking, these objects can be placed in one of two general categories. As Michael McKenna (2008) characterizes them, *hard-line replies* to the four-case argument claim that Plum is morally responsible in all of the manipulation cases, or at the very least that it is not clear that Plum is not responsible in these cases (91). *Soft-line replies*, on the other hand, claim that Plum is responsible in some of the cases although not in others. Mele and Fischer put forth soft-line replies to the four-case argument, while McKenna and Haas opt for hard-line replies. Since the exchange between Fischer and Pereboom elsewhere in this issue addresses soft-line replies, I will focus my attention here on McKenna's hard-line reply and Pereboom's response.

McKenna's central idea is that whatever attitude it is rational initially to have about Plum's responsibility in the ordinary deterministic example transfers to the manipulation cases. As Pereboom describes the objection, "In his view, since it is at the outset rational for us to have an agnostic attitude about the claim that Plum is morally responsible in the ordinary deterministic example..., the absence of relevant differences [between the cases] allows this rational agnosticism to transfer unimpeded to the manipulation cases, thereby depriving them of counting in favor of incompatibilism" (91). According to McKenna, the initial attitude we should have in the ordinary deterministic example is that, "It is not clear that Plum is not morally responsible" (92). Pereboom, however, responds that McKenna is mistaken about the rationality of this initial attitude. As he describes it:

I have a different take on the dialectic...In everyday life, we assume that people can be, and often are, morally responsible in the basic desert sense for their actions. However, we ordinarily do not bring to bear on this assumption any theory about the general causal nature of the universe that might threaten its rationality. For example, we do not seriously question the rationality of this assumption given the theory that every event, including choices and actions, results from deterministic causal processes that trace back to a time before agents existed. (92)

Pereboom maintains that if we did engage in such questioning, the epistemically rational attitude to adopt would be the *neutral inquiring attitude* rather than McKenna's *confirmed agnostic attitude*.

There are a number of different initial attitudes one can bring to the ordinary deterministic example. As Pereboom identifies them, the *resolute compatibilist* response is to "deny that under these circumstances causal determination poses even a prima facie threat to our everyday assumption, and that it is rational to refuse to take seriously any further consideration for there being such a threat" (93). A distinct approach affirms that causal determination provides a reason for giving up the responsibility assumption, but claims that so far the issue has not been settled. Pereboom calls this the *neutral inquiring* response. "By this response it is initially epistemically rational not to believe that the agent in an ordinary deterministic example is morally responsible in the basic desert sense, and not to believe that he isn't, but to be open to clarifying considerations that would make one or other of these beliefs rational" (93). It is crucial to note that the neutral inquiring attitude differs significantly from what Pereboom calls the *confirmed agnostic*. The confirmed agnostic "claims that it is not clear that the ordinary causally determined agent is morally responsible in the sense at issue, and that it is not clear that he isn't, but, like the resolute compatibilist, maintains that it is rational to consider enquiry into the issue closed, and for this reason it is not open to further clarifying considerations" (93).

Pereboom argues that the confirmed agnostic response, the response that generates McKenna's conclusion, is not the appropriate response to take. Instead, "the most attractive way of conceiving manipulation arguments involves supposing that the neutral inquiring attitude about ordinary determined agents is initially epistemically rational" (94). If we adopt the neutral inquiring attitude, "it might then be that an analogous manipulation case functions as a clarifying consideration that makes rational the belief that the ordinary causally determined agent is not morally responsible" (94). While the confirmed agnostic rules out this possibility, the neutral inquirer leaves it open. This means that "the neutral inquiring response is open to the potential rational influence of manipulation examples, and so we cannot assume that it transfers to the manipulation cases unaltered" (94). Pereboom therefore concludes that an argument that

begins with the neutral inquiring response to the ordinary deterministic case (the most epistemically rational attitude to adopt) will not secure agnosticism about manipulation cases.

## Chapter 5: Free Will Skepticism and Rational Deliberation

The remaining four chapters explore the implications of free will skepticism, addresses several practical concerns, and defends the optimistic view that life without free will and desert-based moral responsibility would not be devastating to our conceptions of agency, morality, and meaning in life. Unlike the objections to libertarianism and compatibilism, resistance to free will skepticism is typically driven by practical concerns over its implications: What would it mean for our interpersonal relationships, society, morality, meaning, and the law? What would it do to our standing as human beings? Could we live with the belief that it is true? Critics contend that free will skepticism threatens to “undercut our self-conception as deliberative and rational agents, to make morality incoherent, leave no reason to be moral, render unjustifiable our policies for dealing with wrongdoers, and undermine the emotions and attitudes that lie at the core of human interpersonal relationships” (104). In these concluding chapters Pereboom addresses these concerns one-by-one, and in the process develops the most thoroughgoing and nuanced account of free will skepticism in the literature.

Chapter 5 begins by addressing the concern that free will skepticism, or more generally the belief that our actions are causally determined by factors beyond our control, threatens to conflict with the presuppositions of rational deliberation. Whenever we deliberate about what to do, we typically believe in the “openness” of our options—that is, we believe we have more than one distinct option for which action to perform, each of which is available to us in the sense that we can or could perform each of these actions. It is often argued that belief in such openness is required for deliberation, or at least for rational deliberation (105).

Some have argued, however, that this belief in openness would conflict with the truth of determinism in the sense that, in any deliberative situation, the truth of determinism would rule out the availability to us of all but one distinct option for what to do, and this would rule out openness about what to do (106). The con-

cern is that “a belief required for rational deliberation would be inconsistent with an evident consequence of determinism for one’s actions, and if determinism were true, such a belief would be false” (106). Worse still, “a rational deliberator who believed determinism and its evident consequences for her actions would have inconsistent beliefs” (106). This line of reasoning suggests an incompatibility between rational deliberation and believing in causal determinism—a position Pereboom calls deliberation-incompatibilism.

*Deliberation-incompatibilism:* S’s deliberating and being rational is incompatible with S’s believing that her actions are causally determined (by causal antecedents beyond her control). (106)

The contrary position is:

*Deliberation-compatibilism:* S’s deliberating and being rational is compatible with S’s believing that her actions are causally determined. (106)

Pereboom sets out in ch.5 to develop and defend a version of deliberation-compatibilism.

Taking as its starting point extant compatibilist responses to this challenge, Pereboom begins by rejecting a metaphysical reading of openness, opting instead for an epistemic openness requirement. While other philosophers have developed similar epistemic conditions, existing proposals of this kind have met with significant opposition (107). According to Pereboom, “the preferable position has it that there are two such compatibilist beliefs—or more precisely, compatibilist epistemic states, and that this dual proposal meets what are in effect two distinct strands in the incompatibilist objections to such proposals. One of these specifies an epistemic notion of openness for what to do and the other is an epistemic condition on the efficacy of deliberation” (107). Pereboom’s account of deliberation-compatibilism therefore requires two kinds of epistemic conditions: an *epistemic openness condition* and a *deliberative efficacy condition*.

Pereboom’s epistemic openness condition builds off conditions first proposed by Tomis Kapitan (1986) and Dana Nelkin (2004b). In an attempt to avoid objections to these conditions, Pereboom puts forth the following revised version:

(S) In order to deliberate rationally among distinct actions  $A_1 \dots A_n$ , for each  $A_i$ , S cannot be certain of the proposition that she will do  $A_i$ , nor of the proposition that she will not do  $A_i$ ; and either (a) the proposition that she will do  $A_i$  is consistent with every proposition that, in the present context, is settled for her, or (b) if it is inconsistent with some such proposition, she cannot believe that it is. (113)

Clause (a) deals with cases of consistent *settled* propositions, where a proposition's being settled for an agent is defined as follows:

(Settled) A proposition is settled for an agent just in case she believes it and disregards any uncertainty she has that it is true, e.g., for the purpose of deliberation. (113).

Clause (b) is included to deal with the concern that I can rationally deliberate about whether to do A even if in fact doing A is inconsistent with a proposition I regard as settled in that context. Here, Pereboom maintains, "it is crucial that I then not believe that it is inconsistent; if I did believe this, it's intuitive that I couldn't rationally deliberate about whether to do A" (114).

While condition (S) together with (Settled) is capable of dealing with a number of deliberation-incompatibilist objections, Pereboom maintains that an additional condition is required that incorporates the belief in the efficacy of deliberation. This condition maintains:

(DE) In order to rationally deliberate about whether to do  $A_1$  or  $A_2$ , where  $A_1$  and  $A_2$  are distinct actions, an agent must believe that if as a result of her deliberating about whether to do  $A_1$  or  $A_2$  she were to judge that it would be best to do  $A_1$ , then, under normal conditions, she would also, on the basis of this deliberation, do  $A_1$ ; and similarly for  $A_2$ . (118-19).

This condition is needed to deal with cases like Peter van Inwagen's locked door example:

...imagine that [an agent] is in a room with two doors and that he believes one of the doors to be unlocked and other door to be locked and impassable, though he has no idea which

is which; let him then attempt to imagine himself deliberating about which door to leave by. (1983, 154)

In this example, I am neither certain that I will open door 1, nor that I will not (and the same for door 2). And my opening door 1 (or door 2) is consistent with what is settled for me in the sense specified. "Thus this example poses a threat to (S) together with (Settled) as a compatibilist account for beliefs required for deliberation" (116). It is for this reason that Pereboom adds the efficacy of deliberation requirement. Once (DE) is introduced, van Inwagen's concern, along with several other concerns Pereboom considers, can be addressed since (DE) is not met by the agent in such situations.

The conclusion of ch.5 is that rational deliberation requires satisfaction of both an epistemic-openness condition and a belief-in-deliberative-efficacy condition. According to Pereboom, "(S) together with (Settled), and (DE) appear not to be vulnerable to objections that have been raised against other compatibilist proposals for the beliefs required for deliberation, and this in turn provides reason to think that a deliberation-incompatibilism...can be successfully resisted" (126). If Pereboom's account succeeds, then there is no inconsistency between rational deliberation and believing in causal determinism—that is, it is possible for an agent to believe that causal antecedents beyond her control causally determine her actions and still be capable of rational deliberation.

## Chapter 6: Moral Responsibility without Basic Desert

As we've seen, free will skepticism denies that we have the sort of free will required for moral responsibility in the *basic desert* sense. There are, however, other notions of moral responsibility besides basic desert whose essential features can be endorsed by the skeptic. Pereboom, for example, sets out and defends a forward-looking account of moral responsibility that is perfectly consistent with free will skepticism. He argues that "moral axiological judgments can serve as a foundation for this forward-looking sense of responsibility, and that such judgments are not threatened by the kinds of considerations that undermine moral responsibility in the basic desert sense" (127). While deontological judgments may not fare as well, Pereboom contends that "the axiological moral judgments

and the forward-looking sense of responsibility are sufficient given the interests of morality” (127).

The justification and goal of Pereboom’s forward-looking account of moral responsibility is to “moderate or eliminate dispositions to misconduct” (131) and to provide protection from harm and reconciliation in relationships. It’s reasonable to assume that if such dispositions to misconduct are manifest in past actions they will persist unless corrective measures are taken.

Thus when an agent has acted badly, one might ask him: “Why did you decide to do that? Or, Do you think it was the right thing to do?” where the point of asking such questions is to have him recognize and acknowledge a disposition to behave immorally. If the reasons given in response to such questions indicate that he does have such a disposition, it then becomes apt to request an effort to eliminate it. Engaging in such interactions will be legitimate in light of how they contribute to the agent’s moral improvement. (132)

This model is a variety of the *answerability* sense of moral responsibility defended by Scanlon (1998) and Hilary Bok (1998), and it’s available for free will skeptics to endorse without inconsistency.

Pereboom’s more specific proposal takes as its inspiration a revised version of Michael McKenna’s (2012) recently developed *conversational* theory of moral responsibility. McKenna’s account holds that blaming someone is an expression of a sentiment or attitude, such as moral resentment or indignation, and its function is to communicate moral disapproval to the agent in light of the quality of will she indicated through her actions. As Pereboom describes McKenna’s account:

When a morally responsible person acts, she understands that members of the moral community might assign such a meaning to her action. When her acts are morally charged, she appreciates that she might be introducing a meaningful contribution to such a conversational exchange with others. This initial stage McKenna calls *moral contribution*. The second stage, in which that agent is blamed by a respondent, he labels *moral address*. In the third stage, *moral account*, the blamed agent extends

the conversation by offering an excuse, a justification, or an apology. The respondent might at this point continue the conversation perhaps by forgiving or punishing. In a further stage the blamed agent may be restored to full status in the moral community. This *moral responsibility exchange* is modeled on analogy with an ordinary conversational exchange between speakers of a natural language. (132)

Pereboom maintains that an amended version of McKenna’s account is available for free will skeptics to adopt without inconsistency, as long as it avoids an appeal to basic desert and to expressions of reactive attitudes that are linked with beliefs about basic desert (133).

Pereboom’s modified account grounds blame, not in basic desert, but in three non-desert invoking moral desiderata: “protection of potential victims, reconciliation to relationships both personal and with the moral community more generally, and moral formation” (134). Pereboom argues, “Immoral actions are often harmful, and we have a right to protect ourselves and others from those who are disposed to behave harmfully. Immoral actions can also impair relationships and we have a moral interest undoing such impairment through reconciliation. And because we value morally good character and resulting action, we have a stake in the formation of moral character when it is plagued by dispositions to misconduct” (134). A forward-looking account of moral responsibility grounded in future protection, future reconciliation, and future moral formation is in harmony with free will skepticism since it does not appeal to basic desert in any way.

While Pereboom focuses primarily on his account of blame, he does note that a corresponding account of praise is available as well. Of the three aims of blame—moral formation, protection, and reconciliation—the one most clearly amendable to praise, he argues, is moral formation. “We praise an agent for a morally exemplary action to encourage him to strengthen the disposition that produced it.” Furthermore, “This can have a protective function, since strengthening such dispositions has the effect of reducing dangerous behavior.” As for reconciliation, the corresponding notion with regard to praise could be “celebrating success in a relationship” (135).

On the conversational model, it is the agent's rationality that is engaged in the proposed process for both blame and praise. In the case of blame, for example, "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 he has in fact so acted without excuse or justification, we aim for him to come to see that the disposition issuing in the action is best eliminated" (135). In normal cases, this change is produced by way of the agent's recognition of moral reasons to eliminate the disposition. Accordingly, on Pereboom's account "it is an agent's responsiveness to reasons (cf. Fischer and Ravizza 1998), together with the fact that we have a moral interest in our protection, his moral formation, and our reconciliation with him, that explains why he is an appropriate recipient of blame in this forward-looking sense" (135-36). A similar case can be made for praise. So while many compatibilists see some type of attunement to reasons as the key condition for basic desert moral responsibility, Pereboom instead views it "as the most significant condition for a notion of responsibility that focuses on protection, reconciliation, and moral formation" (136).

A critic may question what form blame will take absent overt expressions of moral resentment and indignation, and will it be effective? But here Pereboom argues, "first, that there are alternative attitudes expression of which is not linked with beliefs about basic desert, communication of which can be as effective morally as expressions of resentment and indignation, and second, that in certain important respects blame without expression of these reactive attitudes is to be preferred (Pereboom 2001, 2009a; cf. Honderich 1988)" (146). For example, when someone is mistreated in a relationship, there are other emotions available besides resentment and indignation—these emotions include "feeling disappointed, hurt or shocked about what the offender has done, moral concern for him, and moral sadness and sorrow generated by this concern when the harm done is serious" (146). Communicating such disappointment, sadness, or concern can be quite effective in motivating avoidance of future misbehavior. In addition, communication of such alternatives to resentment and indignation "is not typically aggressive in the way that expression of anger can be, and will usually not have its intimidating effect" (147).

Pereboom acknowledges that sometimes circum-

stances require a more aggressive and intimidating emotional attitude. But here too skeptical alternatives are available. Take, for example, the right of self defense and defense of others. Here we are justified in expressing fury since it is an attitude consistent with free will skepticism. Fury, according to Pereboom, "is an emotion we share with bears and wolves, and has no cognitive content or presupposition or associated belief that involves the notion of desert" (147). Pereboom's account therefore maintains that "moral sadness and sorrow—accompanied by a resolve for fairness and justice, or to improve personal relationships—will serve societal and personal relationships as well as resent and indignation does. And when acting on the right of self defense and defense of others in a dangerous situation is at issue, fury is appropriate, and not in conflict with free will skepticism" (148).

Additionally, adopting this skeptical account of moral responsibility has the potential benefit of relinquishing us of an often-destructive form of moral anger. Anger in relationships is nourished by the belief that its target is blameworthy in some way for having done wrong (149). Furthermore, the "anger that fuels ethnic conflict often results partly from the belief that an opposing group so deserves blame for some atrocity" (149). Free will skepticism "advocates retracting such beliefs because they are false, as a result of which the associated anger might be diminished, and its expressions reduced" (150). Expression of resentment and indignation is potentially harmful since it is "more likely to occasion destructive resistance than is expression of moral concern and the moral sadness or sorrow that such concern is apt to generate" (150). Furthermore, "overt blame fueled by resentment and indignation arguably renders it particularly susceptible to errors that threaten to undermine the integrity and effectiveness of the moral conversation" (150). It would be beneficial, therefore, if we could rid ourselves of these potentially harmful reactive attitudes, and adopting free will skepticism and Pereboom's forward-looking account of moral responsibility provides us with a path forward for doing so.

## Chapter 7: Free Will Skepticism and Criminal Behavior

Chapter 7 deals with another frequently voiced criticism of free will skepticism: 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” (153). The second, which Pereboom willingly acknowledges, is that “alternative justifications that are not ruled out by the skeptical view *per se* face significant independent moral objections.” Yet despite these concerns, Pereboom forcefully argues that “free will skepticism leaves intact other ways to respond to criminal behavior, in particular preventive detention, rehabilitation, and alteration of relevant social conditions,” and “these methods are both morally justifiable and sufficient for good social policy” (153).

First, it’s important to note 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 (157). For the retributivist, it is the basic desert attached to the criminal’s immoral action alone that provides the justification for punishment. This means that the retributivist position is not reducible to consequentialist considerations nor does it appeal “to a good such as the safety of society, or the moral improvement of the criminal in justifying punishment” (157). 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” (157).

Some critics worry that without retributive punishment the free will skeptic is left unable to adequately deal with criminal behavior. But Pereboom notes 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 Pereboom maintains the first two approaches face independent moral objections—objections that, though perhaps not devastating, make them less desirable than their alternative—he argues 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 Pereboom’s positive account, let me 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” (161). 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. Pereboom notes, however, that 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” (161). Children and adult criminals differ in significant respects. For example, “[a]dult criminals, unlike children, typically understand the moral code accepted in their society” (161). Furthermore, “[c]hildren are generally more psychologically malleable than adult criminals are” (162). For these and other reasons, Pereboom sees this approach as less desirable than an alternative incapacitation account.

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” (163). 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). The most significant pleasure or happiness that results from punishment derives from the security of those who benefit from its capacity to deter. (163–64)

While deterrence theories are completely compatible with free will skepticism, Pereboom notes three general moral objections against them. The first is that they

will justify punishments that are intuitively too severe: “For it would seem that in certain cases harsh punishment would be more effective deterrents than milder forms, while the harsh punishments are intuitively too severe to be fair” (164). The second concern is that such accounts would seem to justify punishing the innocent: “If 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” (164). Lastly, 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” (165).

Despite Pereboom’s rejection of these alternatives to retributivism, he does believe there is 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. As Pereboom describes it:

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 sense for having contracted these diseases. We generally agree that it is sometimes permissible to quarantine them nevertheless. But then, even if a dangerous criminal is not morally responsible for his crimes in the basic desert sense (perhaps because no one is ever in this way morally responsible) it could be as legitimate to preventatively detain him as to quarantine the non-responsible carrier of a serious communicable disease. (156)

Furthermore, this analogy places several constraints on the treatment of criminals.

...as less dangerous diseases justify only preventative measures less restrictive than quarantine, so less dangerous criminal tendencies justify only more moderate restraints. In addition, the incapacitation account that results from this analogy demands a degree of con-

cern 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: 56-9). 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. (156)

This is Pereboom’s incapacitation account and it provides a more resilient proposal for justifying treatment of criminals than either the moral education or deterrence theories of criminal punishment.

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. Concerns over the “use” objection, for example, “count more heavily against punishment policy justified on consequentialist grounds than they do against incapacitation based on the quarantine analogy” (169). And this is because, “on the quarantine analogy, 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” (169). 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” (170). In fact, for certain minor crimes “perhaps only some degree of monitoring could be defended” (170).

Summarizing Pereboom’s proposal, 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” (174). 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” (174). Pereboom’s account also “demands a certain

level of care and attention to the well-being of criminals, which would change much of current policy” (174). 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” (174). This combined approach to dealing with criminal behavior, it is argued, is sufficient for dealing with dangerous criminals, leads to a more humane and affective social policy, and is actually preferable to the harsh and often excessive forms of punishment that typically come with retributivism.

## Chapter 8: Personal Relationships and Meaning in Life

Another set of concerns people have with free will skepticism is that it would negatively impact our personal relationships and meaning in life. Some fear that life without free will would have no point or purpose and a dispirited resignation to fate would result (175). Another closely related concern is that taking the skeptical view would threaten our personal relationships and the fulfillment in life that they provide. P. F. Strawson (1962), for example, developed a philosophical elaboration of this reaction. “For him a skeptical conviction, supposing it was psychologically possible for us, would undermine expressions of the other-directed reactive attitudes essential to good personal relationships, and would jeopardize self-directed reactive attitudes such as guilt, and repentance, crucial to relationships and also to personal moral development” (175). Pereboom, in his final chapter, responds to these concerns and makes his final case for optimistic skepticism.

Pereboom argues that “the skeptical perspective does not threaten personal relationships, and that it holds out the promise of better relationships through release from reactive attitudes such as moral resentment and indignation” (175). Just as the Stoics argued that “affirming determinism while taking a broader perspective can produce an advantageous sort of equanimity” (175), Pereboom maintains that “Spinoza was right to contend that skepticism about free will can encourage a kind of peace of mind that would be a significant benefit for us” (176). While it is an empirical question (currently without answer) whether living in accordance with the skeptical perspective would be on the whole better for us, Pereboom is “optimistic that this

sort of life would be better for us all things considered” (176).

It has already been argued in previous chapters that free will skepticism leaves intact much of what is important to us in human life: the rationality of deliberation, a forward-looking sense of moral responsibility, preventing crime, etc. Here Pereboom considers the additional question: Is the assumption that we are morally responsible in the basic desert sense required for the meaningful and fulfilling personal relationships we have? P. F. Strawson delivers a positive answer, fearing that if we embraced the skeptical perspective we would end up adopting an “objectivity of attitude” toward others, an attitude that rules out good personal relationships.

Pereboom disagrees.

I think that Strawson is right to believe that objectivity of attitude would jeopardize our personal relationships, but that he is mistaken to hold that such a stance would result or be appropriate if the causal determination of our actions by factors beyond our control did pose a genuine threat to expression of the reactive attitudes (Pereboom 1995, 2001). First, some of our reactive attitudes, although their expressions would be doxastically irrational for free will skeptics, are not required for good personal relationships. On the skeptical view, an expression of resentment or indignation will invoke doxastic irrationality when it is accompanied by the belief—as in my view it always is—that its target deserves in the basic sense to be its recipient. But I maintain that expressions of these reactive attitudes are suboptimal as modes of communication in relationships relative to alternative attitudes available to us. Second, the attitudes whose expression we would want to retain either are not threatened by a skeptical conviction because they are not associated with beliefs that conflict with this view, or else they have analogues not connected with such beliefs. The attitudes and analogues expressions of which would survive do not amount to Strawson’s objectivity of attitude, and are sufficient to sustain good personal relationships (Pereboom 1995, 2001, 2007a, 2009a). (179)

The remainder of ch.8 is dedicated to examining the suboptimal nature of resentment and indignation, spelling out the benefits of alternative attitudes and analogues that are left unaffected by the skeptical perspective, and briefly discussing the consequences of free will skepticism for our life's projects.

*Moral resentment*, as Pereboom defines it, is “anger with an agent due to a wrong he has done to oneself,” and *indignation* is “anger with an agent because of a wrong he has done to a third party” (179). While these attitudes play an important communicative role in personal and societal relationships, Pereboom argues that there are other emotions present or available not challenged by the skeptical view that are preferable to anger and “whose expression can also convey the relevant information” (180). These emotions include “feeling hurt or shocked or disappointed about what the offending agent has done, and moral sadness or sorrow and concern for him” (180). These substitutes are preferable to resentment and indignation because the latter attitudes are “apt to have harmful effects.” Expression of resentment and indignation “often fails to contribute to the well being of those to whom it is directed” (180). And as Pereboom notes: “Frequently it is intended to cause physical or emotional pain, and can give rise to destructive resistance instead of reconciliation. As a result, it has the potential to damage or destroy relationships” (180).

To be clear, though, Pereboom is not committed to the view that we can generally succeed in *completely* overcoming moral resentment and indignation, “but rather to the proposal that we can resist such attitudes and limit their expressions with some success, and that we can oppose actions and policies justified on the basis of the beliefs about basic desert that accompany such attitudes” (182). Following Shaun Nichols (2007), one could distinguish between narrow-profile emotional responses, which are local or immediate emotional reaction to situations, and wide-profile responses, which are not immediate and can involve rational reflections. Pereboom maintains that “[f]ree will skeptics can expect that we will not keep ourselves from some degree of narrow-profile, immediate resentment when we are seriously wronged in our most intimate personal relationships.” Nevertheless, “in some wide-profile cases, we could well have the ability to diminish or even eliminate resentment and indignation, or at least disavow it in the sense of rejecting any force it might be thought to have in justifying harmful reactions to

the wrong done, and given a skeptical conviction we might take such measures for the sake of morality and rationality” (181).

Moving on to the attitudes of guilt and repentance, Pereboom addresses the concern that such self-directed attitudes are threatened by free will skepticism. One may object that since these attitudes are essential to good interpersonal relationships for agent like us who can behavior immorally, and since these attitudes are incompatible with free will skepticism, adopting the skeptical perspective would have a negative affect on our relationships. “Without the attitudes of guilt and repentance,” it could be argued, “we would not be motivated to moral improvement after acting badly, we would be barred from subsequent restoration of moral integrity” (186). Pereboom responds, however, by arguing that there are substitutes that work just as well as guilt and repentance.

[S]uppose that you do wrong, but because you believe that free will skepticism is true, you reject the claim that you are blameworthy in the basic desert sense. Instead, you accept that you have done wrong, you feel deeply disappointed that you were the agent of wrongdoing, and as Waller advocates, you feel deep sorrow and regret for what you have done...In addition, because you have commitment to doing what is right, and to personal improvement, you resolve to do what you can to eliminate your disposition to act this way, and perhaps seek to help this change. (186-87)

None of this, argues Pereboom, is undercut by the skeptical conviction.

Similar arguments are given for the attitudes of forgiveness, gratitude, and love. Pereboom argues that the skeptical perspective has no trouble preserving all three of these attitudes—or at least the core aspects of each. Forgiveness, for example, need not presuppose that the person being forgiven deserves to be blamed in the basic desert sense (188). Gratitude, likewise, need not involve the belief that an agent is praiseworthy for some action. For example, “one can be thankful to a young child for some kindness without believing that she is morally responsible for it” (190). And the joy involved in gratitude is also consistent with skepticism: “No feature of the skeptical view poses a threat to the legitimacy of being joyful and expressing

joy when others are considerate or generous in one's behalf" (190). The same can be said about love. "Love of another involves, most fundamentally, wishing well for the other, taking on aims and projects of the other as one's own, and a desire to be together with the other." According to Pereboom, "[f]ree will skepticism does not threaten any of this" (190).

The final concern Pereboom addresses has to do with our life's projects. He asks: "Would it be difficult for us to cope without a conception of ourselves as credit- or praiseworthy for achieving what makes our lives fulfilled, happy, satisfactory, or worthwhile—for realizing what Honderich calls our *life-hopes* (Honderich 1988: 382ff.)?" (193). Pereboom argues that while there is an aspect of these life-hopes that may be undercut by skepticism, the skeptical perspective nevertheless leaves them largely intact. Achievement and life-hopes are not as closely connected to basic desert praiseworthiness as some critics suppose. For example, "[i]f someone hopes for success in some project, and if she accomplishes what she hoped for, intuitively this outcome would be an achievement of hers even if she is not in this particular way praiseworthy for it..." (193).

Free will skepticism need not instill in us an attitude of resignation to whatever our behavioral dispositions together with environmental conditions hold in store. "Even if what we know about our dispositions and environment give us reason to believe that our futures will turn out in a particular way, it can often be reasonable to hope that they will turn out differently" (194). But for this to be so, "it may be important that we lack complete knowledge of our dispositions and environmental conditions" (194). Suppose, for example, that someone reasonably believes that he has a particular disposition that might well be a hindrance to realizing a life-hope. Because he does not know "whether this disposition will in fact have this effect, it remains open for him—that is, epistemically possible for him—that another disposition of his will allow him to transcend this impediment" (194). As a result, he might reasonably hope that he will overcome his fear (or whatever the disposition may be) and achieve his goal. So for the free will skeptic, if he in fact does overcome his fear and succeed at his life's-hope, "this will not be an achievement of his in quite as robust a sense as we might naturally suppose, but it will be an achievement in a substantial sense nonetheless" (194).

For Pereboom, our sense of self-worth "is to a non-trivial extent due to features not produced by our volitions, let alone by free will" (194). He points out that people "place great value on natural beauty, native athletic ability, and intelligence, none of which have their source in our volition" (194). Of course we also value voluntary efforts—"in productive work and altruistic behavior, and in the formation of moral character" (194)—but Pereboom argues that it does not matter much to us that these voluntary efforts are also freely willed. Consider how good character comes to be.

It is plausibly formed to a significant degree by upbringing, and the belief that this is so is widespread. Parents regard themselves as having failed in raising their children if they turn out with immoral dispositions, and they typically take great care to bring their children up to prevent such an outcome. Accordingly, people often come to believe that they have the good moral character they do largely because they were raised with love and skill. But those who believe this about themselves seldom experience dismay because of it. We tend not to become dispirited upon coming to understand that good moral character is not our own doing, and that we do not deserve a great deal of praise or credit for it. By contrast, we often feel fortunate and thankful. (195)

If Pereboom is correct, then there is no reason to think that our sense that we have value and that our lives are worth living is threatened by free will skepticism.

## Conclusion

Pereboom's *Free Will, Agency, and Meaning in Life* makes a strong case for the conclusion that we lack the sort of free will required for moral responsibility in the basic desert sense. It also makes a strong case for optimistic skepticism, arguing that free will skepticism can preserve most of what we care about—including the rationality of deliberation, a forward-looking sense of moral responsibility, preventing crime, and meaning in life. Not only is the skeptical perspective "compatible with a veridical sense of accomplishment when we succeed in our projects," Pereboom argues that it's also compatible with good personal relationships and meaning in life. With regard to our friendships, for example, "it holds out the promise of greater

equanimity by reducing the expression of resentment and indignation that often impairs them” (199). It is for these reasons that Pereboom concludes: “If we did give up the assumption of the sort of free will at issue, then, perhaps surprisingly, we might be better off as a result” (199).

## References

- Ayer, A. J. 1954. Freedom and necessity. In *Philosophical essays*, A.J. Ayer, 271-84. London: Macmillan. Reprinted in *Free will*, ed. Derk Pereboom, 110-18. Cambridge: Hackett Publishing Company, 1997.
- Balaguer, Mark. 2009. *Free will as an open scientific problem*. Cambridge, MA: MIT Press.
- Bentham, Jeremy. 1823. *An introduction to the principles of morals and legislation*. New York: Macmillan.
- Bok, Hilary. 1998. *Freedom and responsibility*. Princeton: Princeton University Press.
- Chisholm, Roderick. 1964. Human freedom and the self. The Lindley Lecture, University of Kansas. Reprinted in *Free will*, ed. Gary Watson, 24-35. New York: Oxford University Press, 1982.
- ———. 1976. *Person and object*. La Salle, IL: Open Court.
- Clarke, Randolph. 1993. Toward a credible agent-causal account of free will. *Nous* 27: 191-203.
- ———. 2003. *Libertarian theories of free will*. New York: Oxford University Press.
- ———. 2005. On an argument for the impossibility of moral responsibility. *Midwest Studies in Philosophy* 29:13-24.
- D'Angelo, Edward. 1968. *The problem of free will and determinism*. Columbia, MO: University of Missouri Press.
- Ekstrom, Laura. 2000. *Free will: A philosophical study*. Boulder, CO: Westview.
- Feinberg, Joel. 1970. Justice and personal desert. In *Doing and deserving: Essays in the theory of responsibility*, ed. Joel Feinberg. Princeton: Princeton University Press.
- Fischer, John Martin. 1982. Responsibility and control. *Journal of Philosophy* 79:24-40.
- ———. 1994. *The metaphysics of free will: A study of control*. Malden, MA: Blackwell Publishing.
- ———. 2007. Compatibilism (and) Responses to Kane, Pereboom, and Vargas. In *Four views on free will*, Robert Kane, John Martin Fischer, Derk Pereboom, and Manuel Vargas, 44-84, 184-90. Oxford: Blackwell.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Franklin, Christopher. 2011. Neo-Frankfurtians and buffer cases: The new challenge to the principle of alternative possibilities. *Philosophical Studies* 152: 189-207.
- Frankfurt, Harry G. 1969. Alternative possibilities and moral responsibility. *Journal of Philosophy* 66 (23): 829-39.
- ———. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1): 5-20.
- Ginet, Carl. 1990. *On action*. Cambridge: Cambridge University Press.
- ———. 1996. In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Philosophical Perspectives* 10: 403-17.
- ———. 2002. Review of *Living without free will*. *Journal of Ethics* 6: 305-09.
- ———. 2007. An action can be both uncaused and up to the agent. In *Intentionality, deliberation, and autonomy*, ed. C. Lumer and S. Nannini, 243-56. Farnham, UK: Ashgate.
- Goetz, Stewart. 2008. *Freedom, teleology, and evil*. London: Continuum.
- Griffith, Meghan. 2010. Why agent-caused actions are not lucky. *American Philosophical Quarterly* 47:43-56.
- Haji, Ishtiyaque. 2004. Active control, agent causation, and free action. *Philosophical Explorations* 7:131-48.
- Hobbes, Thomas. 1654. *Of libertie and necessity: A treatise, wherein all controversie concerning predestination, election, free-will, grace, merits, reprobation, etc. is fully decided and cleared, in answer to a treatise written by the Bishop of London-Derry, on the same subject*, London, printed by W. B. for F. Eaglesfield, 1654.
- Honderich, Ted. 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*, 1990.
- Hume, David. 1748/2000. *An enquiry concerning human understanding*, ed. L. A. Selby-Bigge. New York: Oxford University Press.
- Hunt, David. 2000. Moral responsibility and unavoidable action. *Philosophical Studies* 97: 195-227.

- ———. 2005. Moral responsibility and buffered alternatives. *Midwest Studies in Philosophy* 29:126-45.
- Kane, Robert. 1985. *Free will and values*. Albany: SUNY Press.
- ———. 1996. *The significance of free will*. New York: Oxford University Press.
- ———. 2007. Libertarianism (and) Responses to Fischer, Pereboom, and Vargas. In *Four views on free will*, Robert Kane, John Martin Fischer, Derk Pereboom, and Manuel Vargas. Oxford: Blackwell.
- Kant, Immanuel. 1781/1787/1987. *Critique of pure reason*. Tran. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.
- Kapitan, Tomis. 1986. Deliberation and the presumption of open alternatives. *Philosophical Quarterly* 36: 230-51.
- McCann, Hugh. 1998. *The works of agency*. Ithaca: Cornell University Press.
- McKenna, Michael. 2005. Where Frankfurt and Strawson meet. *Midwest Studies in Philosophy* 29:163-80.
- ———. 2008. A hard-line reply to Pereboom's four-case argument. *Philosophy and Phenomenological Research* 77:142-59.
- ———. 2012. *Conversation and responsibility*. New York: Oxford University Press.
- Mele, Alfred R. 1999. Ultimate responsibility and dumb luck, *Social Philosophy and Policy* 16: 274-93.
- ———. 2005. A critique of Pereboom's 'four-case argument' for incompatibilism. *Analysis* 65:75-80.
- ———. 2006. *Free will and luck*. New York: Oxford University Press.
- Moore, G. E. 1912. Free will. In *Ethics*, Moore, 84-95. Oxford: Oxford University Press.
- Moya, Carlos. 2006. *Moral responsibility: The ways of skepticism*. Oxford University Press.
- Nelkin, Dana. 2004b. Deliberative alternatives. *Philosophical Topics* 32: 215-40.
- ———. 2011. *Making sense of freedom and responsibility*. Oxford: Oxford University Press.
- Nichols, Shaun. 2007. After incompatibilism: A naturalistic defense of the reactive attitudes. *Philosophical Perspectives* 21:405-28.
- O'Connor, Timothy. 2000. *Persons and causes*. New York: Oxford University Press.
- ———. 2003. Review of *Living without free will*. *Philosophical Quarterly* 53:308-10.
- ———. 2009. Agent-causal power. In *Dispositions and causes*, ed. Toby Handfield, 189-214. Oxford: Clarendon Press.
- Palmer, David. 2011. Pereboom on the Frankfurt cases. *Philosophical Studies* 153:261-72.
- Pereboom, Derk. 1995. Determinism *al dente*. *Noûs* 29:21-45.
- ———. 2000. Alternative possibilities and causal histories. *Philosophical Perspectives* 14: 119-37.
- ———. 2001. *Living without free will*. New York: Cambridge University Press.
- ———. 2003. Source incompatibilism and alternative possibilities. In *Freedom, responsibility, and agency: Essays on the importance of alternative possibilities*, eds. M. McKenna and D. Widerker, 185-99. Aldershot, UK: Ashgate Press: 185-99.
- ———. 2004. Is our conception of agent causation incoherent? *Philosophical Topics* 32:275-86.
- ———. 2007a. Hard incompatibilism (and) Responses to Kane, Fischer, and Vargas. In *Four views on free will*, Robert Kane, John Martin Fischer, Derk Pereboom, and Manuel Vargas, 85-125, 191-203. Oxford: Blackwell.
- ———. 2007b. On Mele's *Free will and luck*. *Philosophical Explorations* 10:163-72.
- ———. 2009a. Free will, love, and anger. *Ideas y Valores: Revista de Colombiana de Filosofía* 141:5-25.
- ———. 2009b. Further thoughts about a Frankfurt-style argument. *Philosophical Explorations* 12: 109-18.
- ———. 2014. *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Reid, Thomas. 1788/1983. Essays on the active powers of man. In *The works of Thomas Reid*, ed. Sir William Hamilton. Hildesheim: G. Olms Verlagsbuchhandlung.
- Sartorio, Carolina. 2011. Actuality and responsibility. *Mind* 120:1071-97.
- ———. 2013. Making a difference in a deterministic world. *Philosophical Review* 122:189-214.
- Scanlon, Thomas. 1998. *What we owe to each other*. Cambridge: Harvard University Press.
- ———. 2013. Giving desert its due. *Philosophical Explorations* 16:101-16.
- Shabo, Seth. 2010. Uncompromising source incompatibilism. *Philosophy and Phenomenological Research* 80: 349-83.
- Strawson, Galen. 1986. *Freedom and belief*. Oxford: Oxford University Press.
- ———. 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.
- Stump, Eleonore. 1990. Intellect, will, and the principle of alternative possibilities. In *Christian theism and the problems of philosophy*, ed. Michael Bealy. Notre Dame, IN: University of Notre Dame Press: 254-85.
  - ———. 1996. Libertarian freedom and the principle of alternative possibilities. In *Faith, freedom, and responsibility*, eds. Jeff Jordan and Daniel Howard Snyder. Lanham, MD: Rowman and Littlefield: 73-88.
  - Taylor, Richard. 1966. *Action and purpose*. Englewood Cliffs, NJ: Prentice-Hall.
  - ———. 1974. *Metaphysics*. 4<sup>th</sup> ed. Englewood Cliffs, NJ: Prentice-Hall.
  - Van Inwagen, Peter. 1983. *An essay on free will*. Oxford: Clarendon Press.
  - Vihvelin, Kadri. 2013. *Causes, laws, and free will*. New York: Oxford University Press.
  - Wallace, R. Jay. 1994. *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.
  - Widerker, David. 1995. Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *The Philosophical Review* 104: 247-61.
  - ———. 2006. Libertarianism and the philosophical significance of Frankfurt scenarios. *Journal of Philosophy* 103: 163-97.
  - Zagzebski, Linda. 1991. *The dilemma of freedom and foreknowledge*. New York: Oxford University Press.
  - ———. 2000. Does libertarian freedom require alternative possibilities? *Philosophical Perspectives* 14: 231-48.