

Free Will

In *The Oxford Handbook of the History of Ethics*, Roger Crisp, ed., Oxford: Oxford University Press, 2013, pp. 606-37.

Derk Pereboom, Cornell University

Introduction

The problem of free will and moral responsibility arises from a conflict between two powerful considerations. On the one hand, we human beings typically believe that we are in control of our actions in a particularly weighty sense, very different from how a thermostat controls a furnace's operation, or a computer controls the marking of typographical errors. We express this sense of difference when we attribute moral responsibility to human beings but not to such machines. Traditionally, we suppose that moral responsibility requires us to have some type of free will in producing our actions, and hence we assume that humans, by contrast with the thermostat and the computer, have this sort of free will. At the same time, there are reasons for regarding human beings as relevantly more like mechanical devices than we ordinarily imagine. These reasons stem from various sources: most prominently, from scientific views that consider human beings to be components of nature and therefore governed by natural laws, whether deterministic or statistical, and from theological concerns that require everything that occurs to be causally determined by God. For many contemporary philosophers, the first of these is especially compelling, and as a result, they accept causal determinism or else claims about the universe similarly threatening to moral responsibility.

The history of philosophy records three standard types of reaction to this dilemma. *Compatibilists* maintain that it is possible for us to have the free will required for moral

responsibility if determinism is true. Others contend that determinism is not compossible with our having the free will required for moral responsibility -- they are *incompatibilists* -- but they resist the reasons for determinism and claim that we do possess this free will of this kind. They advocate the *libertarian* position. *Hard determinists* are also incompatibilists, but they accept that determinism is true and that we lack the sort of free will required for moral responsibility.

Especially since David Hume's discussion of these issues (Hume 1739/1978, 1748/1999; cf. Russell 1995), the concern about the existence of the sort of free will required for moral responsibility has been extended to whether it is compatible with the *indeterminacy* of actions. This development has challenged the value of the threefold classification just canvassed, despite its persistence in the contemporary debate. In particular, some maintain that the free will required for moral responsibility is not only incompatible with determinism, but with at least some varieties of indeterminism as well. Agent-causal libertarians typically hold that this type of free will is incompatible with the kind of indeterminism according to which only events (or states, or property-instances) are causes. Free-will skeptics typically agree. A skeptic such as Galen Strawson (1986, 1994) maintains that this kind of free will is incompatible with any variety of indeterminism. I argue that it is incompatible only with the event-causal sort, and not with indeterministic agent causation, but that agent-causation is empirically implausible (Pereboom 1995, 2001). Complications arise on the compatibilist side as well. Hume, and later R. E. Hobart (1934) and Alfred Ayer (1954), contend that while the sort of free will required for moral responsibility is compatible with determinism, it is in fact incompatible with indeterminism, at least with indeterminism located at the point at which a decision or an

intention is produced. This tradition continues in the work of philosophers such as Ishtiyaque Haji (1998) and Alfred Mele (2006). A different sort of compatibilism, according to which this sort of free will is compatible with both determinism and indeterminism, is inspired by some remarks of Hume's, and is developed in detail by P. F. Strawson in "Freedom and Resentment" (Strawson 1962). In this view, the practice of holding people morally responsible has its own internal system of norms, but is not subject to a legitimate external challenge, from, for example, general scientific discoveries about the universe. Whether the universe is deterministic or indeterministic is claimed to be irrelevant to whether our holding agents morally responsible is legitimate, and in this respect Strawson's compatibilism is *insulationist*.

It is important to recognize that the term 'moral responsibility' is used in a number of ways, and that the moral responsibility in several of these senses is uncontroversially compatible with the causal determination of action by factors beyond our control. Yet there is one particular sense of moral responsibility, and a correlative type of free will, that have been at play in the historical debate, and these notions are not uncontroversially compatible with this sort of determinism:

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

Basic-desert entailing moral responsibility is typically thought to be presupposed by our retributive reactive attitudes, such as indignation and moral resentment. In P. F. Strawson's (1962) account, moral responsibility is essentially tied to these reactive attitudes, and hence the basic-desert entailing sense is plausibly the variety that he brings to the fore.

Alternative notions of moral responsibility have not been a focus of the free will debate. For example, an agent could be considered morally responsible just in case it is legitimate to expect her to respond to such questions as: "Why did you decide to do that? Do you think it was the right thing to do?" and that she evaluate critically what her actions indicate about her moral character. Engaging in such interactions might well be reasonable in light of the way in which they contribute to our own and others' moral improvement (Scanlon 1998, Bok 1998). However, incompatibilists would not regard the control required for morally responsibility in such an "answerability" sense is to be incompatible with determinism. The type of moral responsibility that incompatibilists do claim to generate an incompatibility with determinism is instead characterized by basic desert and the reactive attitudes that presuppose it. From this point on, unless otherwise indicated, I will use the term 'moral responsibility' to refer to this particular variety.

Source and leeway theories

In recent decades, by contrast with the previous history of the debate, it has been common to claim that agent's moral responsibility for an action is not explained (at least not primarily) by the availability to her of alternative possibilities, for example by the ability to do otherwise than what she has actually done. Rather, responsibility is to be explained by the agent's being the actual source of her action in a specific way. One might

thus adopt a *source* as opposed to *leeway* position on the sort of free will required for moral responsibility. Examples of the kind devised by Harry Frankfurt supply the standard challenge to the leeway position (Frankfurt 1969). An agent considers performing some action, but an intervener is concerned that she will not come through. So if she were to show some sign that she will not or might not perform the action, the intervener would arrange matters so that she would perform it anyway. Here is one of John Fischer's examples: Jones will decide to kill Smith only if Jones blushes beforehand. Jones's failure to blush (by a certain time) can then function as the prior sign that would occasion the intervention that would cause her to kill Smith. Suppose that Jones acts without the intervention. We might well have the intuition that she is morally responsible for killing Smith, although she could not have done otherwise than to kill Smith, and despite the fact that she could not even have formed an alternative intention. More cautiously, Fischer concludes that if Jones is not morally responsible for his action, it wouldn't be because he couldn't have done otherwise. Jones could have failed to blush, but Fischer contends that such a "flicker of freedom" is of no use to the libertarian, since it is not *robust* enough to play a part in grounding her moral responsibility (Fischer 1994, 131-59).

Frankfurt-style arguments have met with significant opposition (e.g., Kane 1985; Widerker 1995, 2006; Ginet 1997; McKenna 1997, Otsuka 1998; Wyma 1998; Franklin 2011; Nelkin 2011; Palmer 2011), and Frankfurt-defenders have devised cases intended to avoid the concerns raised (Stump 1996; Mele and Robb 1998; Pereboom 2000, 2001, 2009, 2011; Hunt 2000, 2005; McKenna 2003; Widerker 2006; Fischer 2010). If there are indeed successful cases, both compatibilist and incompatibilist versions of the source position emerge as options, and each of these types of view has its advocates. According to source

incompatibilism, moral responsibility requires that the agent be the source of her action in a way incompatible with her being causally determined to act by factors beyond her control. It might well be that alternative possibilities -- not necessarily of the robust sort -- are entailed by her being the source of her action in this way. But these alternative possibilities would not have the primary role in explaining an agent's moral responsibility. Rather, they would be a consequence of the factor that did: the agent's being the source of her action in the right way (Della Rocca 1998, Pereboom 2001).

For proponents of the leeway position, both compatibilist and incompatibilist, the accessibility of alternative possibilities is crucial for explaining why an agent would be morally responsible. For them, despite the claims of the defenders of Frankfurt examples, the intuition that an agent's moral responsibility for an action requires that she could have done otherwise retains considerable force. This force is aptly captured by what David Widerker calls the *W-defense*. About an agent who breaks a promise, but could not have done otherwise he writes:

Still, since you, [Harry] Frankfurt, wish to hold him blameworthy for his decision to break his promise, tell me *what, in your opinion, should he have done instead?* Now, you cannot claim that he should not have decided to break the promise, since this was something that was not in his power to do. Hence, I do not see how you can hold Jones blameworthy for his decision to break the promise. (Widerker 2000: 191)

Michael McKenna expresses the source theorists' main response to the W-Defense. When Widerker asks of an agent, in view of the fact that he had no robust alternative possibility: what would you have him do? it should be admitted that there is no good answer. Yet in opposition to this initially disturbing result, we should instead draw attention to what the

agent has actually done, and to the actual causal history by which his action came about (McKenna 2005).

Compatibilism

Retaining the legitimacy of our ordinary attitudes towards human actions and at the same time regarding them as causally determined has been so attractive that a large proportion of contemporary philosophers classify themselves as compatibilists. Two prominent routes to compatibilism can be distinguished. The first and more common kind 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. This route to compatibilism is typically developed by surveying our intuitions about blameworthiness and praiseworthiness in specific kinds of examples -- involving, for instance, coercion, addiction, mental illness, hypnotism, and brainwashing. These reactions are then employed to motivate conditions on causal integration required for moral responsibility. Compatibilisms of this sort have arguably been advanced by Aristotle (cf. Irwin 1990; Meyer 1993; the Stoics (cf. Inwood 1985; Bobzien 1998; Brennan 2005); Hobbes (1654; cf. Pink 1994), and Hume (1739/1978, 1748/1999; cf. Russell 1995), and in the 20th century by G. E. Moore (1912), R. E. Hobart (1934), A. J. Ayer (1954), Harry Frankfurt (1971), Daniel Dennett (1984, 2003), Gary Watson (1975), John Fischer (1994), with Mark Ravizza (1998), Jay Wallace (1994), Ishtiyaque Haji (1998), Alfred Mele (2006), and at least with respect to praiseworthiness, by Susan Wolf (1990) and Dana Nelkin (2011)). The second route to compatibilism is the

one advocated by P. F. Strawson, which 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. In his view, the basis of moral responsibility is to be found in reactive attitudes such as indignation, moral resentment, guilt, and gratitude. For example, the fact that agents are typically resented for certain kinds of immoral actions is what constitutes their being blameworthy for performing them. In particular, justification for claims of blameworthiness and praiseworthiness ends in the system of human reactive attitudes. Because moral responsibility has this type of basis, the truth or falsity of determinism is immaterial to whether we are justified in holding agents morally responsible.

The Humean tradition, developed with intensity during the first half of the twentieth century by philosophers such as Hobart (1934) and Ayer (1954), features a family of causal integrationist conditions. Perhaps most prominently, Hume himself advocates a conditional specification of the ability to act otherwise. To be free in the sense at issue, an agent must satisfy the following criterion: “by liberty, then, we can only mean a power of acting or not acting according to the determinations of the will – that is, *if we choose to remain at rest, we may; if we choose to also move, we also may*” (1748/1999, section 8, my emphasis). G.E. Moore (1912) advanced a view of this type, arguing that to say that I could have acted otherwise is to claim that I would have acted otherwise *if I had so chosen*. Even if I am causally determined to act as I do, it might still be true that I would have acted otherwise if I had chosen so to act. This position was highly prominent over the subsequent half-century; Ayer (1954), for example, provides a classic statement of this view.

Such conditional analyses of ‘could have done otherwise’ were challenged by C. D.

Broad (1934), C. Campbell (1951), and especially forcefully in the 1960's by Roderick Chisholm (1964) and Keith Lehrer (1968) by way of the following type of consideration. Suppose Brown does not at some time t jump in the sea to save a drowning child, and we say:

(1) Brown could have jumped into the sea at t .

A proponent of conditional analysis proposes that (1) is equivalent to:

(2) If Brown had chosen to jump into the sea at t , he would have jumped into the sea at t .

(Variants of (2) substitute for 'chosen': 'willed' 'tried' 'set himself' or 'wanted.')

Chisholm and Lehrer argue that this sort of analysis is subject to a counterexample of the following form. Suppose that the sea is very cold, and Brown knows it, and as a result it is psychologically impossible for him to choose to jump into the sea. But we might suppose that if he did choose to jump, he would actually jump. Thus here (2) is true, and yet it is strongly intuitive that (1) is false – Brown could not have jumped into the sea. The conclusion is that (2) is not a correct analysis of (1). Recent times, however, have seen the development of more sophisticated conditional analyses (Vihvelin 1991, 2004; Fara 2008); for a critical review, see Clarke (2009)).

Second, Humeans propose that *absence of constraint* is an important necessary condition for moral responsibility. In the *Treatise of Human Nature* Hume characterizes *liberty of spontaneity* as "that which is opposed to violence" (1739/1978: 407), arguably intending that an action is performed with liberty of spontaneity just in case the agent is not constrained to act as she does. In Ayer's (1954) exposition, when an agent is under constraint, desires that genuinely belong to her do not play the causal role necessary for

her action to be free in the sense required for moral responsibility. Against this it is frequently argued that despite initial appearances, absence of constraint is not a necessary condition for moral responsibility. Suppose someone threatens to kill Green unless he kills five other people, and Green proceeds to kill them. Green could nevertheless be morally responsible for this action, even though he was constrained to act as he does. Moreover, one could be morally responsible even if one acts under certain internal sorts of constraint. Kleptomania is often cited as a paradigm of such a condition (Ayer 1954). But whether kleptomania undermines responsibility depends on certain characteristics of this illness. Perhaps if it merely strongly inclines an agent towards stealing, and he can nevertheless resist the inclination, this condition does not undermine moral responsibility.

A third member of the Humean family is the claim that an agent is morally responsible for an action only if it flows from her "durable and constant" character (Hume 1739/1978: 411). One difficulty for this proposal is that an agent who kills his superior for firing him might then be exempt from moral responsibility if he is at all other times law-abiding and not given to violent behavior. It would appear that in some cases of this sort, the action will be out of character, but intuitively the agent could still be morally responsible.

Harry Frankfurt (1971) sets out another highly influential causal integrationist route to compatibilism. On his proposal, an action is free in the sense required for moral responsibility when the first-order desire that results in action conforms in a certain way to the agent's second-order desires. More precisely,

1. *First-order desires* are identified by statements of the form "A wants to X" in which the term "to X" refers to an action.

2. The desire identified by "A wants to X" is (part of) A's will just in case "A wants to X" is either the desire by which he is motivated in some action he performs or the desire by which he will or would be motivated when or if he acts. The will consists in *effective desires*, as opposed to, for example, desires that one has that never would result in action.

3. *Second-order desires* are identified by statements of the form "A wants to X," in which the term "to X" refers to a first-order desire.

4. A *second-order volition* is a kind of second-order desire, and is identified by a statement of the form "A wants to want X" when it is used to mean that A wants X to be part of his will, i.e. he wants to will X (and not that A wants merely to want X without willing X).

5. A person *acts freely and of his own free will* just in case he wills X and wants to will X (and, Frankfurt specifies, he wills X *because* he wants to will X – although this clause is typically ignored).

Acting freely and of one's own free will is the type of freedom required for moral responsibility. Suppose that Scarlet kills Mustard, and this is what Scarlet desired to do, and she wanted her desire to kill Mustard to be effective. According to Frankfurt, Scarlet then satisfies the crucial condition for acting freely in the sense required for moral responsibility. Frankfurt can be viewed as attempting to capture the intuition that for an agent to be morally responsible for an action is for it to be really hers, for it genuinely to belong to her. What it is for an action to genuinely belong to an agent is for her to *identify* with it, and the mechanism of identification is a second-order desire for a first-order willing, that is, a second-order volition.

From the development of Stoicism onward (Inwood 1985; Bobzien 1998; Brennan 2005), it has often been argued that the feature of human agency that undergirds moral responsibility is the capacity to rationally regulate one's actions. When we take people to be the sorts of beings who are morally responsible, we expect them to govern their behavior by, for example, choosing on the basis of available reasons, considering good reasons that were previously ignored, and weighing heretofore unappreciated reasons differently. In recent times, John Fischer and Mark Ravizza (1994, 1998), Jay Wallace (1994), Susan Wolf (1990), and Dana Nelkin (2011) have developed positions in which a capacity to regulate one's behavior by reasons is the crucial condition for moral responsibility.

Fischer initially argued that a fairly weak notion of *reasons-responsiveness* is what is crucially required for moral responsibility:

An agent is *weakly reasons-responsive* when a certain kind K of mechanism, which involves the agent's rational consideration of reasons relevant to the situation, issues in action, and in *at least* some alternative circumstances in which there are sufficient reasons for her to do otherwise than she actually does, she would be receptive to these reasons and would have chosen and done otherwise by the efficacy of the same deliberative mechanism that actually results in the action. (1994: 166-7).

By 'sufficient reason' Fischer means 'justificatorily sufficient reason,' that is, a reason that is, all things considered, an agent's strongest or best reason for action. Thus White is morally responsible when she decides to pay her telephone bill, for the usual reasons, next week rather than today, if, in circumstances in which she knew that her telephone would be

disconnected if she did not pay today -- and she has the ordinary sufficient reasons to want her telephone be functional -- she would, by the deliberative mechanism that actually results in her deciding to pay next week, appreciate the different reasons and decide to pay today instead. If her practical reasoning would not differ in varying circumstances in which there are sufficient reasons to do otherwise, she would not be morally responsible.

In response to counterexamples in which the agent is weakly reasons-responsive and yet acts in accord with a pattern of reasons so odd that it seems clear that the agent is not morally responsible, Fischer and Ravizza (1998) later required that the mechanism by which the agent acts be receptive to a *pattern* of reasons that is intuitively rational. They call the revised criterion *moderate reasons-responsiveness*. But at the same time, they in effect continued to endorse the original weak version of *reasons-reactivity*, on the ground that agents can be morally responsible when they are receptive to sufficient reasons to do otherwise, but yet do not react to those reasons (1998: 65-82). In addition, they contend that to be morally responsible, the agent must, through a historical process, *take responsibility* for the springs of her action, and this involves coming to see herself as a rational agent, and as a fair target of the reactive attitudes, in a way that is appropriately based on evidence.

Like Fischer and Ravizza's proposal, Wallace's focuses on our capacity for regulating our behavior by reasons (Wallace 1994). First of all, in his view, *holding a person morally responsible* is to be characterized as applying to particular acts, with essential reference to the reactive attitudes of resentment, indignation, and guilt, and to moral obligations that one, as moral judge, accepts. What is required for legitimately holding an agent morally responsible for a particular act is that it would be appropriate to have a reactive attitude of

indignation, resentment, or guilt directed toward the agent in that situation, and what would make such an attitude appropriate is the agent's having violated a moral obligation (1994: 58-93). Wallace distinguishes the conditions of holding someone morally responsible, which apply to a particular act, from the conditions of *accountability*, which apply to agents over an extended period. To regard someone as a morally accountable agent is "to view the person as the sort of agent whose violation of moral obligations one accepts would render reactive emotions appropriate" (1994: 70). In order for an agent to be legitimately considered morally accountable, she must possess the *powers of reflective self-control*: the power to grasp and apply moral reasons, and the power to control or regulate her behavior by the light of such reasons (1994: 154-94).

Psychopaths, for example, are not morally accountable because they lack these powers of reflective self-control, as is the case for those who have a strong form of drug addiction (1994: 170-8). Yet normal agents, even given the truth of causal determinism, do have these powers. It is important for Wallace that an agent may possess the powers for reflective self-control even though a particular circumstance prevents him from applying these powers. In his terminology, the requirement of having these powers is thus to be understood *generally*. In making this point, he aims to preempt the incompatibilist criticism that supposing the murderer was causally determined to commit his crime, he did not have the capacity to regulate his behavior by moral reasons when he acted (1994: 207-21). Wallace maintains that even if the murderer did not have the capacity to apply the powers in this particular case he could still be morally responsible because he retains the powers in the general sense.

Susan Wolf (1990) contends that an agent is morally responsible just in case she

could have acted in accordance with Reason -- that is, just in case she could have done the right thing for the right reasons. On her conception, this condition results in an asymmetry, on the assumption that an agent can never do otherwise due to causal determinism. For if the agent is causally determined to do the right thing for the right reasons, then a fortiori she can do this, and hence she meets the proposed condition on responsibility. But if she is causally determined to do the wrong thing for the wrong reasons, then it might well be that she could not have done the right thing for the right reasons, whereupon she would not be morally responsible. Recently, Nelkin (2011) provides a development of this position in which this asymmetry is further grounded in a parallel asymmetry regarding the "ought implies can" principle. By this principle, if one does not do what one ought, then one must have been able to do otherwise, but if one does what one ought, this ability is not required. Given plausible connections to praiseworthiness and blameworthiness and praiseworthiness, the result is that blameworthiness requires the ability to do otherwise, but praiseworthiness does not.

The second type of route to compatibilism is suggested by Hume in a discussion of the effect the thesis of divine determinism does and should have on the moral sentiments (1748/1999, section 8). He first argues for a psychological thesis, that the sentiments of approbation and blame cannot be affected by a belief in theological determinism or in any philosophical theory; and subsequently for a normative thesis, that these sentiments should not be affected by a belief in theological determinism or in any philosophical theory. Because the generation of sentiments of approbation and blame certain circumstances is inevitable and not preventable by us, having these sentiments is compatible with any general metaphysical claim about the nature of reality and should not be controlled or

altered by any philosophical theory (cf. Russell 1995).

P. F. Strawson's (1962) position has a similar core, but is developed in more detail; (for related views and relevant discussions, see Watson 1987, Arpaly 2006, and McKenna 2011). Strawson begins by characterizing two opposing participants in the controversy about determinism and moral responsibility, the "optimist" and the "pessimist." The optimist is a compatibilist of the sort who justifies the practices of moral disapproval and punishment on the basis of their social utility (e.g., Schick 1939). The pessimist is an incompatibilist who wants libertarianism to be true. Strawson rejects the incompatibilism and libertarianism of the pessimist, but he also contends, with the pessimist, that the optimist is missing something significant in the account of moral responsibility.

Rather than justifying the practice of holding people morally responsible solely on utilitarian grounds, in Strawson's account it is the reactive attitudes to which people are subject by virtue of participation in interpersonal relationships that play the central role. He proposes that these reactive attitudes provide what is missing in the optimist's theory. In fact, in his view it is the reactive attitudes alone that provide the foundation for our holding people morally responsible. To secure his case for compatibilism, Strawson, like Hume, argues first for a psychological thesis, that the reactive attitudes cannot be affected by a belief in causal determinism, and then for a normative thesis, that they should not be affected by this belief. Thus, what fills the role of what the pessimist correctly judges to be missing in the optimist's account of our moral life -- our reactive attitudes -- cannot and should not be undermined by a belief in causal determinism.

The reactive attitudes are natural human reactions to the good or ill will of others towards us, as displayed in their attitudes and actions. Strawson's examples include

gratitude, moral resentment and indignation, guilt, forgiveness, repentance and love. These attitudes are essential features of ordinary human interpersonal relationships. Sometimes we can and at times we appropriately do forgo or suspend reactive attitudes. Excuses invite us to view an injury as one in respect of which a particular one of these attitudes is inappropriate even though the agent is more generally a fitting target for reactive attitudes. Exemptions, by contrast, invite us to view the agent as someone in respect of whom these attitudes are inappropriate. In Strawson's view this involves adopting an objective attitude towards him, which amounts to treating impersonally, as something to be controlled or managed. But none of this shows that a belief in universal determinism would lead to a suspension of reactive attitudes, for "the human commitment to participation in ordinary interpersonal relationships is... too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships as we normally understand them; and being involved in interpersonal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question." According to Strawson, both the optimist and the pessimist are wrong in thinking that the reactive attitudes must be justified from outside of the practice of holding morally people morally responsible in which these attitudes are embedded. This is, in fact, where their deepest mistake lies. The framework of these attitudes "neither calls for, nor permits, an external 'rational' justification." (1962: 70)

A number of criticisms have been raised against Strawson's position. A first is that he is mistaken to hold that the reactive attitudes cannot be affected by the belief that determinism is true. Another is that supposing a belief in determinism did issue in a

theoretical or epistemic conflict with the ordinary reactive attitudes, it is not clear that rational appraisal of these attitudes would always favor retention over revision. A third is that it is mistaken to maintain that a challenge to the reactive attitudes based on causal determinism is external to the practice of holding people morally responsible and therefore illegitimate. Rather, exemptions from moral responsibility that are widely regarded as acceptable and are thus internal to the practice will generalize to an incompatibilist condition on moral responsibility. Manipulation arguments against compatibilism, which we will examine in the next section, require that suitably manipulated agents are not responsible, and that relevantly similar cases should receive the same moral assessment, and each of these principles seems internal to the practice of holding agents morally responsible.

Incompatibilism.

Various arguments have been devised to support the incompatibilist's claim that free will and determinism are not compossible. One prominent contention is that the relevant sort of free will is the ability to do otherwise, and this ability is precluded by determinism. For if determinism is true, then facts about the remote past, together with the laws of nature, entail every subsequent fact, including facts about actions. The Consequence Argument aims to develop this idea with rigor and precision (e.g., Ginet 1966, Lamb 1977, van Inwagen 1983; cf. Warfield 2000).

Peter van Inwagen's version of the Consequence Argument begins with a characterization of determinism:

Determinism =df the non-relational facts of the past (P), indexed to any time, t, in

conjunction with a complete description of the laws of nature (L), entails every nonrelational fact at any time later than t (F).

To this are added two principles expressing metaphysical truths, the *Principle of the Fixity of the Past*, FP, and the *Principle of the Fixity of the Laws*, FL. FP states that for any agent, the past is fixed for her relative to the present; she is powerless to render false any truth about the past (this needs to be made more precise in view of “soft” facts about the past, such as the partition of India’s occurring more than 60 years prior to your choice to continue reading this article). FL asserts that, for any person, the laws of nature are fixed; it’s not up to her whether or not they obtain. She is powerless to render false any law of nature.

At this point we can define a “powerlessness” modal operator ‘Np,’ which can be read as “p is true, and [the relevant agent] is powerless to render p false. One can first of all derive true claims of powerlessness by this inference rule:

Rule α : $\Box p \rightarrow Np$

where ‘ \Box ’ as expresses logical necessity. Applying Rule α , one can conclude that any logically necessary truth is such that any selected agent is powerless to render it false. Next, van Inwagen invokes an inference principle that transfers powerlessness from one fact to another one that is entailed by the first:

Rule β : $N(p \rightarrow q), Np \rightarrow Nq$.

With these resources, the Consequence Argument can now be employed to show that any action performed in a deterministic world is one the agent is powerless to avoid. Let s be the claim that a sample actual action is performed. Then:

1. $\Box(P \cdot L \rightarrow s)$ Assume determinism; apply it to s.
2. $\Box(P \rightarrow (L \rightarrow s))$ 1, and Logic.

- 3. $N(P \rightarrow (L \rightarrow s))$ 2, Rule α .
- 4. $N(P)$ Assume FP.
- 5. $N(L \rightarrow s)$ 3, 4, Rule β .
- 6. $N(L)$ Assume FL.
- 7. $N(s)$ 5, 6, Rule β .

This argument tells us that, at the determined world W , the agent is powerless to render s false; in other words, the agent is powerless to avoid the action referred to in s .

The consequence argument has been strongly contested. Frequently discussed objections include: contrary to 6, the agent can act in such a way that, if she were to so act, the laws of nature that do obtain at W would not (Lewis, 1981); Rule β is invalid and therefore one cannot infer the powerlessness of the agent to avoid her action from the truth of determinism and her lack of power over the past and the laws (McKay and Johnson 1996; Carlsson 2000); and the powerlessness invoked in the argument does not uncontroversially transfer across events that involve agency that satisfies compatibilist conditions on moral responsibility (Slote 1982; Bishop 1989).

On a source compatibilist position, the availability of alternative possibilities does not characterize the crucial kind of free will, and it is thus open to such a compatibilist to contend that the Consequence Argument does not challenge her view. But a cousin of the Consequence Argument, the Direct Argument, also set out by van Inwagen, does pose a threat to source compatibilism (1983). The Direct Argument's core principle transfers *non-responsibility* across entailment relations, instead of powerlessness, to force the conclusion that any action is such that its agent is not responsible for it. The more general proposal is that given the assumption of determinism, non-responsibility transfers from the past and

the laws to all subsequent events. One important objection to this argument is McKenna's, and it is analogous to Slote's response to the Consequence Argument: Non-responsibility does not uncontroversially transfer across contexts in which the compatibilist conditions on moral responsibility are satisfied by agents (McKenna 2008).

Another prominent type of argument against compatibilism begins with the intuition that if an agent is causally determined to act by, for example, scientists who manipulate her brain, then she is not morally responsible for that action, even if she meets compatibilist conditions on moral responsibility (Taylor 1974; Ginet 1990; van Inwagen 1983; Pereboom 1995, 2001; Mele 1995, 2006; Kane 1996). The subsequent step is that there are no differences between the manipulated agents and their ordinary deterministic counterparts that can justify the claim that the manipulated agents are not morally responsible while the determined agents are.

My multiple-case version of such an argument first of all develops examples of an action that results from an appropriate sort of manipulation, and in which the prominent compatibilist conditions on moral responsibility are satisfied (Pereboom 1995, 2001, 2011). These examples, taken separately, indicate that it is possible for an agent not to be morally responsible even if the compatibilist conditions are met, and that as a result these conditions are inadequate. But the argument has additional force by virtue of setting out three such cases, each of which is progressively more like a fourth, in which the action is causally determined in a natural way. The first case involves manipulation that is local and determining, and hence very likely to elicit the non-responsibility intuition. The second is like the first, except it restricts manipulation to the beginning of the agent's life. The third is similar, except the manipulation results from strict community upbringing. The aim is to

formulate the cases so that it is not possible to draw a principled line between any two adjacent cases that would explain why the agent would not be morally responsible in the first but would be in the second, largely because all of the prominent compatibilist conditions are satisfied in each. The conclusion is that the best explanation for why the agent is not responsible in the four cases is that he is causally determined by factors beyond his control in each of them, and this result conflicts with the compatibilist's central claim.

Libertarianism.

Incompatibilism divides into two positions, the libertarian view according to which determinism is false and we have the free will required for moral responsibility, and the hard determinist perspective in which determinism is true and we lack this kind of free will. Although libertarianism has from ancient times been a frequently-held view, a number of its most prominent contemporary features were articulated in the late 13th and 14th centuries by Franciscans Peter Olivi, John Duns Scotus, and William of Ockham, and then in the 16th century by Francisco Suarez. Scotus defends the position that the will is subject to two basic affections, for happiness and for justice, and that the will is free to choose in accordance with either, independently of how the intellect represents the options. While in the spirit of the Platonic tradition Aquinas held that the will is restricted to what the agent's intellect represents as good, Scotus maintained that for each of the two basic affections, the agent is capable of willing what it represents as best, and also as not-willing, or refraining from willing, this option. Thus, in the case of each of these affections the agent has the capacity for alternatives, defined not as the power to will and to positively will otherwise,

but as the power to will and to refrain from willing (Scotus 1986; Adams 1986, 1995).

Suarez's (1597) formulation was highly influential and widely disseminated beginning in the late 16th century. In it the essentials of modern libertarianism are in place: following the Franciscans, his view specifies that agents are free in the sense that by virtue of their will and power they have the ability both to perform and to refrain from performing the action at issue; and he explicitly characterizes of the kind of free will at stake as the sort required for basic-desert-involving moral responsibility (Penner, ms; see also Pink, forthcoming).

More recent times have witnessed the explicit differentiation of two major versions of libertarianism, the event-causal and the agent-causal types. In event-causal libertarianism, actions are caused solely by way of events, standardly conceived as objects having properties at times, and some type of indeterminacy in the production of actions by appropriate events is held to be a decisive requirement for moral responsibility (Kane 1996; Ekstrom 2000). This position has an ancestor in the Epicurean view according to which a free decision is an uncaused swerve in the otherwise downward path of an atom (Lucretius 50 BCE). According to agent-causal libertarianism, free will of the sort required for moral responsibility is accounted for by the existence of agents who possess a causal power to make choices without being determined to do so (Kant 1781/1787/1987, cf. Wood 1984, Watkins 2005, Pereboom 2006; Reid 1788/1983, cf. Rowe 1991, Yaffe 2004; Taylor 1966, 1974; Chisholm 1964, 1976; O'Connor 2000; Clarke 1993, 2003; Griffith 2010). In this view, it is essential that the causation involved in an agent's making a free choice is not reducible to causation among events involving the agent, but is rather irreducibly an instance of the agent-as-substance causing a choice not by way of events. The agent, fundamentally as a substance, has the causal power to cause choices without

being determined to do so.

Robert Kane (1996) has developed a much-discussed version of event-causal libertarianism. In his proposal, the paradigm case of an action for which an agent is responsible is one of moral or prudential struggle, in which there are reasons for and against performing the action in question. The sequence that produces the action begins with the agent's character and motives, and proceeds through the agent's making an effort of will to act, which results in the choice for a particular action. The effort of will is a struggle to choose in one way in a situation in which there are countervailing pressures. In the case of a freely-willed choice, this effort of will is *indeterminate*, and as a result the choice produced by the effort is *undetermined*. Kane explains this last specification by drawing an analogy between an effort of will and a quantum event (on one conception; Kane 1996: 128). The effort of will is indeterminate in the sense that its causal potential does not become determinate until the choice occurs. Prior to this pivotal interaction, there are different ways in which this causal potential can be resolved, and thus when it is resolved, the resulting choice will be undetermined. Significantly, Kane cautions against construing his view in such a way that the indeterminacy occurs after the effort is made: "One must think of the effort and the indeterminism as fused; the effort is indeterminate and the indeterminism is a property of the effort, not something that occurs before or after the effort." If an agent is morally responsible for a choice, it must either be free in this sense or there must be some such free choice that is (or has a key role in) its sufficient ground, cause, or explanation (1999: 232). Kane embellishes his account by endeavoring to show how the particle analogy for free choice might actually work in the functioning of the brain's neural networks (1996: 128-30).

Critics of libertarianism have argued that if actions are undetermined, agents cannot be morally responsible for them. A classical presentation of this objection is found in Hume's *Treatise of Human Nature* (Hume 1739/1978: 411-2; cf. Mele 2006). In Hume's version, the concern highlighted is that if an action is uncaused, it will not have sufficient connection with the agent for her to be morally responsible for it. This idea might profitably be explicated as follows. For an agent to be morally responsible for a decision, she must exercise a certain type and degree of control in making that decision. In an event-causal libertarian picture, the relevant causal conditions antecedent to a decision – agent-involving events -- would leave it open whether this decision will occur, and the agent has no further causal role in determining whether it does. With the causal role of these antecedent conditions already given, it remains open whether the decision occurs, and whether it does is not settled by anything about the agent. So whether the decision occurs or not is in this sense a matter of luck, and, intuitively, the agent lacks the control required for moral responsibility for the decision.

Libertarians agree that an action's resulting from a deterministic sequence of causes that traces back to factors beyond the agent's control would rule out her moral responsibility for it. The deeper point of this objection is that if this sort of causal determination rules out moral responsibility, then it is no remedy simply to provide slack in the causal net by making the causal history of actions indeterministic. Such a move would yield requirement for moral responsibility -- the absence of causal determinism for decision and action -- but it would not supply another — sufficiently enhanced control (Clarke 1997, 2003). In particular, it would not provide the capacity for an agent to be the

source of one's decisions and actions that, according to many incompatibilists, is unavailable in a deterministic framework.

The agent-causal libertarian's solution to this problem is to specify a way in which the agent could have this enhanced control, which involves the power to settle which of the antecedently possible decisions actually occurs. The proposed solution is to reintroduce the agent as a cause, this time not merely as involved in events, but rather fundamentally as a substance (for detractors, see Haji 2004; Mele 2006). The agent-causal libertarian maintains that we possess a distinctive causal power – a power for an agent, fundamentally as a substance, to cause a decision without being causally determined to do so. This position might well be incipient in the medieval originators of modern libertarianism, but the Humean objection to indeterministic free will occasioned a more precise formulation. Thomas Reid provides an agent causal account in direct response to Hume (1788/1983), and arguably, Kant does as well in his account of transcendental freedom (Kant 1781/1787/1987, cf. Watkins 2005; Reid 1788/1983; Chisholm 1964, 1976; O'Connor 2000; Clarke 2003).

One traditional objection to the agent-causal picture is that we have no evidence that we are substances of the requisite sort. Kant expresses two further concerns for the agent-causal view, which in his account calls for our endorsement on practical, but not on evidential grounds. One is that despite our inability to discern an internal incoherence in our conception of transcendental freedom, our having this power might nonetheless not be really or metaphysically possible. In Kant's view, we can discern something to be really possible only through experience, and we do not experience ourselves as agent causes. The second concern is that agent causation might not be reconcilable with what we would

expect given our best empirical theories. Kant himself believed that the physical world, as part of the world of appearance, is governed by deterministic laws, while the transcendently free agent-cause would exist not as an appearance, but as a thing in itself. In this 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 encounter divergences from the deterministic laws. For the alterations in the physical world that result from the undetermined decision would themselves not be causally determined, and they would thus not be governed by deterministic laws. One might object that it is possible that the physical alterations that result from every free decision just happen to dovetail with what could in principle be predicted on the basis of the deterministic laws, so nothing actually occurs that diverges from these laws. However, this proposal would seem to 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. Kant thought that more needed to be said. In one text he appears to claim that when an agent makes a transcendently free decision, he, as atemporal noumenal subject, also freely produces everything in the past that causally determines his free actions (1788/1996: Ak 97-8). But this sort of atemporalist line is at best insignificantly more credible than an overt contradiction (Wood 1984; Pereboom 2006).

More recent expositors of the agent-causal view, such as Randolph Clarke (1993, 2003) and Timothy O'Connor (2000, 2009), suggest that quantum indeterminacy can help

with the reconciliation project. 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. Suppose, as is controversial, that significant quantum indeterminacy percolates up to the level of human action. Then it might seem that agent-causal libertarianism can be reconciled with the claim that the laws of physics govern at least the physical components of human actions. However, it appears that wild coincidences would also arise on this suggestion (Pereboom 1995, 2001). Consider the class of possible actions each of which has a physical component whose antecedent probability of occurring is approximately 0.32. It would not violate the statistical laws in the sense of being logically incompatible with them if, for a large number of instances, the physical components in this class were not actually realized close to 32% of the time. Rather, the force of the statistical law is that for a large number of instances it is correct to *expect* physical components in this class to be realized close to 32% of the time. Are free choices on the agent-causal libertarian model compatible with what the statistical law would have us to expect about them? If they were, then for a large enough number of instances the possible actions in our class would almost certainly be freely chosen near to 32% 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% of the time would also amount to a wild coincidence. The proposal that agent-caused free choices do not diverge from what the statistical laws predict for the physical components of our actions would be so sharply opposed to what we would expect as to make it incredible.

At this point the agent-causal libertarian might propose that exercises of agent-causal libertarian freedom do result in divergences from what we would expect given our

best assessments of the physical laws. Roderick Chisholm (1964) proposes such a view. Divergences from the probabilities that we would expect absent agent-causes do in fact occur whenever we act freely, and these divergences are located at the interface between the agent-cause and that part of the physical world that it directly affects, likely to be found in the brain. Chisholm remarks that whenever an agent acts freely, a miracle occurs. There are different ways in which agent-caused free choices might diverge from what the laws would predict. One way is by not being subject to laws at all. Another is by being subject to different statistical laws, an option on which the agent-cause would be governed by probabilistic laws that are its own due to the fact that they emerge only in the right sorts of agential contexts, and not to those that generally govern the physical events of the universe (O'Connor 2008). A concern for these kinds of claims is that we currently have no evidence that they are true.

A third type of libertarianism is non-causal, and is advocated by Henri Bergson (1889/1910), Carl Ginet (1990), Hugh McCann (1998), and Stewart Goetz (2008). Bergson argues that although action occurs in time, the time of conscious agency does not divide into the kinds of quantities or magnitudes that make objects in space subject to causal natural laws. On Ginet's conception, the key conditions for a basic action's being free are that it has an agent as a subject, it has an actish phenomenological feel, and it is uncaused, or not causally determined. An often-cited objection to this type of libertarianism is that control in action is fundamentally a causal matter, and thus such theories cannot secure the type of control in action required for moral responsibility (O'Connor 2000, Clarke 2003). This objection is resisted by non-causalists, and such theories continue to be attractive, especially given the problems that arise on causal libertarian theories due to the

expectation that the causality of libertarian agency must be reconciled with the causality of the natural world.

Skepticism about free will

Due to the difficulties for both the compatibilist and libertarian positions, some have endorsed that skeptical outlook that we do not have the sort of free will required for moral responsibility. Hard determinists argue that because determinism is true and compatibilism is implausible, we lack free will of this kind. This perspective is defended by Baruch Spinoza (1677/1985), Paul Holbach (1770), Joseph Priestley (1788/1965), B. F. Skinner (1970), Ted Honderich (1988), and Bruce Waller (1990). Critics have expressed a number of concerns about this type of view. They have argued, for example, that hard determinism would threaten our self-conception as deliberative agents, that it would undercut the reactive attitudes essential to good human interpersonal relationships, and that if hard determinism were true morality itself would be incoherent.

Hard determinists maintain that it is because of the truth of causal determinism that we lack the sort of free will required for moral responsibility (e.g., Spinoza 1677/1985: 440-4, 483-4, 496-7; cf. Della Rocca 2008). By contrast, many contemporary free will skeptics are agnostic about causal determinism. I argue, for example, that we would not be morally responsible if determinism were true, but also that we would lack moral responsibility if indeterminism was true and the causes of our actions were exclusively events. Such indeterministic causal histories of actions would be as threatening to this sort of free will as deterministic histories are. However, it might be that if we were undetermined agent-causes we would then have this type of free will. But although our

being undetermined agent causes has not been ruled out as a coherent possibility, it is not credible given our best physical theories. Thus I do not claim that our having the sort of free will required for moral responsibility is impossible. Nevertheless, since the only account on which we are likely to have this kind of free will is not credible given our best physical theories, we must take very seriously the prospect that we are in fact not free in the sense required for moral responsibility.

Galen Strawson, also a free-will skeptic, argues that moral responsibility requires a conception of agency that human beings could not satisfy, and its impossibility for us can be established independently of an examination of the truth of determinism (1986: 25-60; 1994). Strawson, then, is a no-free-will-either-way theorist, that is, he maintains that for us moral responsibility is incompatible with both determinism and indeterminism. The specific type of agent-causal libertarianism whose coherence I defend is relevantly similar to what he calls the "Leibnizian" variety of agent causation. In this conception, the agent's reasons (which consist of beliefs and desires) constitute only one component of the cause of a free decision, since they may incline but do not all by themselves cause the agent to choose. The agent's causation of a choice constitutes the remaining part. But, Strawson contends -- as he does against any theory of moral responsible agency -- that rational actions must have a full causal explanation in terms of the agent's reasons alone. More precisely, rational actions must have an explanation in terms of reasons the agent has that indicates all of what there was about the agent, mentally speaking, that causally brought it about that she performed the action she did (1986: 52-6). Otherwise only the "reasons" part of the action's cause would be rational, and the remaining mental aspects would not be. "Upon what," he asks, "are the decisions about actions now supposed to be based, other than upon its reasons?" (1986: 53). The decisions

cannot be based on further principles of choice or on further reasons, because the same questions can be asked about those. Either the reasons all by themselves cause the decision, in which case the agent-cause has no causal role and the decision is not free, or the agent-cause plays a part that is not rational. Consequently, these agent-caused actions must either be unfree or "rationally speaking random," and thus rationally deficient.

But one might contest Strawson's argument against agent causation on the ground that the standard for rationality it assumes is too stringent. Imagine that you are in a situation of conflict between self-interested and moral reasons; one possible decision is morally justified but is not on balance in your self-interest, the other is not morally justified but is in your self-interest. You have the capacity to freely agent-cause either choice, you cause the moral one, and this decision is on balance justified by your reasons overall. Strawson would judge that the agent is unfree because the choice is not fully caused by your reasons, and thus rationally speaking random. Yet in this situation it is intuitive that the action might well be sufficiently connected to your reasons to count as rational. The rationality of action would then not require its complete determination by the agent's reasons, but rather only being justified by them (cf. Pereboom 2001, Clarke 2005).

Saul Smilansky concurs with Strawson's argument that the sort of free will required for moral responsibility is impossible for us, and thus he too is a no-free-will-either-way theorist (1997, 2000). He argues, however, that a valuable type of moral self-respect, respect for oneself as the moral agent one has freely made oneself, would be undermined if we did not believe we are free in the sense required for moral responsibility. Having the sort of self-respect at issue requires maintaining or fostering an *illusion* of free will, which Smilansky recommends for this reason, and for others as well, such as justification of criminal

punishment.

My own view is that the denial of moral responsibility in the basic-desert sense does issue in the deleterious moral consequences it is often thought to have (Pereboom 2001). In criminology, for instance, the free-will skeptic might endorse a theory that draws an analogy between the treatment of criminals and the treatment of carriers of dangerous diseases. Ferdinand Schoeman (1979) argues that if we have the right to quarantine carriers of serious communicable diseases to protect people, then for the same reason we also have the right to isolate the criminally dangerous. Quarantining a person can be justified when he is not morally responsible for being dangerous to others. If a child is infected with a deadly contagious virus transmitted to her before she was born, quarantine can still be justified. By analogy, even if a violent criminal is not morally responsible for his crimes (say because no one is ever morally responsible), it might well be as legitimate to isolate him as it would be to quarantine a non-responsible carrier of a dangerous communicable disease.

On skepticism's consequences for interpersonal relationships, P. F. Strawson may be right that if we persistently maintained an objective attitude toward others, our relationships would be undermined (but see Sommers 2007). But one might reasonably deny that skepticism about free will and moral responsibility would justify adopting this attitude in particular. Indeed, certain reactive attitudes would in fact be threatened, since some of them, such as moral resentment, indignation and a variety of guilt would have the false presupposition that the person who is the object of the attitude is morally responsible in the basic-desert entailing sense. But one might relinquish these attitudes while retaining nonreactive emotions such as moral sorrow and regret, which might have similar salutary

roles in relationships. In addition, the attitudes that we most clearly want to retain, such as gratitude and love, either would not be not threatened in this way, or else have analogues or aspects that would not have false presuppositions. For example, one aspect of gratitude is simply to be thankful to someone for what she has done, and mere thankfulness does not require a presumption of basic desert. Thus it might well be that the surviving emotions would not amount to the objective attitude, and at the same time would be sufficient to sustain good human relationships (Pereboom 2001, 2009).

About the view that we lack the sort of free will at issue, Spinoza says: “this doctrine contributes to the social life insofar as it teaches us to hate no one, to disesteem no one, to mock no one, to be angry at no one...” (1677/1985: 490). Often we justify expressions of moral resentment or indignation by contending that their targets are morally responsible in the basic-desert entailing sense for what they have done. If we became convinced that we lack the type of free will required for this sort of moral responsibility, we would regard such justifications as illegitimate. Spinoza is concerned about the damage to which these reactive attitudes give rise, and in his view our coming to believe that we lack the sort of free will that justifies their expression is on balance to our advantage.

References

- Adams, Marilyn (1986). "The Structure of Ockham's Moral Theory," *Franciscan Studies* 24, pp. 1-35.
- Adams, Marilyn (1995). "Duns Scotus on the Will as Rational Power," *Via Scoti, Methodologica ad Mentem Joannis Dunds Scoti*, ed. Leonardo Sileo (Rome: PAA Edizioni Antonianum, pp. 839-54.
- Aristotle (357/2000). *Nicomachean Ethics*, tr. Roger Crisp, Cambridge: Cambridge University Press.
- Arpaly, Nomy (2006). *Merit, Meaning, and Bondage: An Essay on Free Will*, Princeton NJ: Princeton University Press.
- Ayer, Alfred J. (1954). "Freedom and Necessity," in Alfred J. Ayer, *Philosophical Essays*, London: Macmillan.
- Bergson, Henri (1889/1910). *Essai sur les données immédiates de la conscience*, Paris: F. Alcan, 1889; translated as *Time and Free Will*, tr. F. L. Pogson: London: Allen and Unwin, 1910.
- Bishop, John (1989). *Natural Agency*, Cambridge: Cambridge University Press.
- Bobzien, Suzanne (1998). *Determinism and Freedom in Stoic Philosophy*, Oxford: Oxford University Press.
- Bok, Hilary (1998). *Freedom and Responsibility*, Princeton NJ: Princeton University Press.
- Brennan, Tad (2005). *The Stoic Life: Emotions, Duties, and Fate*, Oxford: Oxford University Press.
- Broad, C. D. (1934) "Determinism, Indeterminism, and Libertarianism," inaugural lecture delivered at Cambridge in 1934; reprinted in C. D. Broad, *Ethics and The History of Philosophy*, London: Routledge, 1952.
- Campbell, C. A. "Is Free Will a Pseudo-Problem?" *Mind* 60, pp. 446-65.
- Carlson, Erik (2000), "Incompatibilism and the Transfer of Power Necessity", *Noûs*, 34: 277-290.
- Chisholm, Roderick (1964). "Human Freedom and the Self," The Lindley Lecture, Department of Philosophy, University of Kansas, 1964; reprinted in *Free Will*, ed. Gary Watson, Oxford: Oxford University Press, 1982.

- Chisholm, Roderick. *Person and Object*, La Salle: Open Court, 1976.
- Clarke, Randolph. (1993). "Toward a Credible Agent-Causal Account of Free Will," *Noûs* 27, pp. 191-203.
- Clarke, Randolph (1997). "On the Possibility of Rational Free Action," *Philosophical Studies* 88, pp. 37-57.
- Clarke, Randolph (2000). "Modest Libertarianism," *Philosophical Perspectives* 14, 2000.
- Clarke, Randolph (2003). *Libertarian Theories of Free Will*, Oxford: Oxford University Press.
- Clarke, Randolph (2005). "On an Argument for the Impossibility of Moral Responsibility," *Midwest Studies in Philosophy* 29, pp. 13-24.
- Clarke, Randolph (2009). "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind* 118, pp. 323-351.
- Della Rocca, Michael (1998). "Frankfurt, Fischer, and Flickers," *Noûs* 32, pp. 99-105.
- Della Rocca, Michael (2008). *Spinoza*, London, Routledge.
- Dennett, Daniel (1984). *Elbow Room*, Cambridge: MIT Press.
- Dennett, Daniel (2003). *Freedom Evolves*, New York: Viking.
- Ekstrom, Laura W. (2000). *Free Will: A Philosophical Study*, Boulder: Westview.
- Fara, Michael (2008). "Masked Abilities and Compatibilism," *Mind* 117, pp. 843-65.
- Fischer, John Martin (1986). "Responsibility and Control," in *Moral Responsibility*, John Martin (Ithaca: Cornell University Press), pp. 174-90.
- Fischer, John Martin (1994). *The Metaphysics of Free Will*, Oxford: Blackwell.
- Fischer, John Martin (2010). "The Frankfurt Cases: The Moral of the Stories," *Philosophical Review* 119, pp. 315-36.
- Fischer, John Martin, and Mark Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*, (New York: Cambridge University Press).
- Fischer, John Martin, Robert Kane, Derk Pereboom, Manuel Vargas (2007). *Four Views on Free Will*, Oxford: Blackwell Publishers.

- Foot, Philippa (1957). "Free Will as Involving Determinism," *The Philosophical Review* 66.
- Frankfurt, Harry G. (1969). "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, pp. 829-39.
- Frankfurt, Harry G. (1971). "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, pp. 5-20.
- Franklin, Christopher (2011a). "Neo-Frankfurtians and Buffer Cases: The New Challenge to the Principle of Alternative Possibilities," *Philosophical Studies* 152, pp. 189–207.
- Ginet, Carl. "Might We Have No Choice," in *Freedom and Determinism*, ed. Keith Lehrer, New York: Random House, 1966, pp. 87-104.
- Ginet, Carl (1996). "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing," *Philosophical Perspectives* 10, pp. 403-17.
- Ginet, Carl (1997). "Freedom, Responsibility, and Agency," *Journal of Ethics* 1, pp. 85-98.
- Goetz, Stewart (2008). *Freedom, Teleology, and Evil*, London: Continuum.
- Griffith, Meghan (2010). "Why Agent-Caused Actions are not Lucky," *American Philosophical Quarterly* 47, pp. 43-56.
- Haji, Ishtiyaque (1998). *Moral Appraisability*, New York: Oxford University Press.
- Haji, Ishtiyaque (2004). "Active Control, Agent Causation, and Free Action," *Philosophical Explorations* 7, pp. 131-48.
- Hobbes, Thomas (1654). *Of Libertie and Necessity. A treatise, wherein all controversie concerning predestination, election, free-will, grace, merits, reprobation, &c., 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
- Holbach, Paul (1770). *Système de la Nature*, Amsterdam.
- Honderich, Ted (1988). *A Theory of Determinism*, Oxford: Oxford University Press.
- Hume, David (1739/1978). *A Treatise of Human Nature*, Oxford: Oxford University Press, 1978.
- Hume, David (1748/1999). *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press, 1999.
- Hunt, David (2000). "Moral Responsibility and Unavoidable Action," *Philosophical Studies*

97, pp. 195-227.

Hunt, David (2005). "Moral Responsibility and Buffered Alternatives," *Midwest Studies in Philosophy* 29, pp. 126-45.

Inwood, Brad (1985). *Ethics and Human Action in Early Stoicism*, Oxford: Oxford University Press.

Irwin, Terence (1990). *Aristotle's First Principles*, Oxford: Oxford University Press.

Kane, Robert (1985). *Free Will and Values*, Albany: SUNY Press.

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

Kane, Robert (1999). "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism," *Journal of Philosophy* 96, pp. 217-40.

Kant, Immanuel (1781/1787/1997). *Critique of Pure Reason*, Paul Guyer and Allen Wood, tr., Cambridge: Cambridge University Press, 1997.

Kant, Immanuel (1788/1996). *Critique of Practical Reason*. Mary Gregor, tr., Cambridge: Cambridge University Press, 1996.

James, William. "The Dilemma of Determinism," in *The Will to Believe and Other Essays*, New York, Longman, 1909.

Lamb, James (1977). "On a Proof of Incompatibilism," *Philosophical Review* 86: pp. 20-35.

Lehrer, Keith (1968). "Cans without Ifs," *Analysis*, 29, pp. 29-32.

Lewis, David (1981). "Are We Free to Break the Laws?", *Theoria*, 47, pp. 113-2

Lucretius (50 BCE). *De Rerum Natura*, translated by W. H. D. Rouse, Loeb Classical Library, Cambridge MA: Harvard University Press, 1982.

McCann, Hugh (1998). *The Works of Agency*, Ithaca: Cornell University Press.

McKay, Thomas and Johnson, David (1996). "A Reconsideration of an Argument against Compatibilism", *Philosophical Topics*, 24: 113-22.

McKenna, Michael (1997). "Alternative Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy* 28, pp. 71-85.

McKenna, Michael (2003). "Robustness, Control, and the Demand for Morally Significant Alternatives," in *Freedom, Responsibility, and Agency: Essays on the Importance of*

- Alternative Possibilities*, Michael McKenna and David Widerker, eds., Aldershot UK: Ashgate, pp. 201-16.
- McKenna, Michael (2005). "Where Frankfurt and Strawson Meet," *Midwest Studies in Philosophy* 29, pp. 163-80.
- McKenna, Michael (2008). "Saying Good-Bye to the Direct Argument in the Right Way," *Philosophical Review* 117 (2008), pp. 349-83.
- McKenna, Michael (2011). *Conversation and Responsibility*, New York: Oxford University Press.
- Mele, Alfred (1995). *Autonomous Agents*, New York: Oxford University Press.
- Mele, Alfred (1999). "Ultimate Responsibility and Dumb Luck," *Social Philosophy and Policy* 16, pp. 274-293.
- Mele, Alfred (2005). "Libertarianism, Luck, and Control," *Pacific Philosophical Quarterly* 86, pp. 395-421.
- Mele, Alfred (2006). *Free Will and Luck*, Oxford: Oxford University Press.
- Mele, Alfred, and David Robb (1998). "Rescuing Frankfurt-Style Cases," *Philosophical Review* 107, pp. 97-112.
- Meyer, Susan Sauvé (1993). *Aristotle on Moral Responsibility*, Oxford: Oxford University Press.
- Moore, G. E. (1912). "Free Will," in G. E. Moore, *Ethics*, Oxford: Oxford University Press.
- Moya, Carlos (2006). *Moral Responsibility: The Ways of Skepticism*, Oxford: Oxford University Press.
- Nelkin, Dana (2008). "Responsibility and Rational Abilities: Defending an Asymmetrical View," *Pacific Philosophical Quarterly* 89, pp. 497-515.
- Nelkin, Dana. (2011). *Making Sense of Freedom and Responsibility*, Oxford: Oxford University Press.
- Nichols, S. (2007). "After Compatibilism: a Naturalistic Defense of the Reactive Attitudes," *Philosophical Perspectives* 21, pp. 405-28.
- O'Connor, Timothy (2000). *Persons and Causes*, New York: Oxford University Press.
- O'Connor, Timothy (2008). "Agent-Causal Power," in *Dispositions and Causes*, ed. T.

- Handfield, Oxford: Oxford University Press.
- Otsuka, Michael (1998). "Incompatibilism and the Avoidability of Blame," *Ethics* 108, pp. 685-701.
- Palmer, David (2011). "Pereboom on the Frankfurt Cases," *Philosophical Studies*, forthcoming
- Pasnau, Robert (1999). "Olivi on Human Freedom," in *Pierre De Jean Olivi (1248–1298)*, Paris: Vrin, pp. 15–25.
- Penner, Sydney (ms). "Suarez on Human Freedom."
- Pereboom, Derk (1995). "Determinism *Al Dente*," *Nous* 29, pp. 21-45.
- Pereboom, Derk (2000). "Alternative Possibilities and Causal Histories," *Philosophical Perspectives* 14, 2000.
- Pereboom, Derk (2001). *Living Without Free Will*, Cambridge: Cambridge University Press.
- Pereboom, Derk (2004). "Is Our Conception of Agent Causation Incoherent?" *Philosophical Topics*. 32, pp. 275-86.
- Pereboom, Derk (2009). "Further Thoughts about a Frankfurt-Style Argument," *Philosophical Explorations* 12 (2010), pp. 109-18.
- Pereboom, Derk. (2010), "Optimistic Skepticism about Free Will," *The Philosophy of Free Will: Selected Contemporary Readings*, eds. Paul Russell and Oisín Deery, New York: Oxford University Press, forthcoming.
- Priestley, Joseph (1788/1965). *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity, in a Correspondence between Dr. Price and Dr. Priestley*; Part III reprinted in Joseph Priestley, *Priestley's Writings on Philosophy, Science, and Politics*, ed. John Passmore, New York: Collier, 1965, pp. 147-52.
- Pink, Thomas (2004). *Free Will: A Very Short Introduction*, New York: Oxford University Press.
- Reid, Thomas (1788/1983). *Essays on the Active Powers of Man*, in *The Works of Thomas Reid, D.D.*, ed. Sir William Hamilton, Hildesheim: G. Olms Verlagsbuchhandlung, 1983.
- Rowe, William (1991). *Thomas Reid on Morality and Freedom*, Ithaca, NY: Cornell University Press.

- Russell, Paul (1995). *Freedom and Moral Sentiment*, New York: Oxford University Press.
- Scotus, John Duns (1986). *Duns Scotus on the Will and Morality*, ed. and tr. Alan Wolter, Washington, DC: The Catholic University of America Press.
- Slote, Michael (1982). "Selective Necessity and the Free Will Problem," *Journal of Philosophy* 79, pp. 5-24.
- Suarez, Francisco (1597). *Disputationes Metaphysicae*, Vivès edition, 28 volumes (1856-1861), Paris; reprinted by Georg Olms, Hildesheim, 1965.
- Scanlon, Thomas (1998). *What We Owe to Each Other*, Cambridge MA: Harvard University Press.
- Schoeman, Ferdinand D. (1979). "On Incapacitating the Dangerous," *American Philosophical Quarterly* 16, pp. 27-35.
- Schlick, Moritz (1939). "When is a Man Responsible," in *Problems of Ethics*, David Rynin, tr., New York: Prentice-Hall, 1939, pp. 143-56.
- Smilansky, Saul (1997). "Can a Determinist Help Herself?," C.H. Manekin and M. Kellner, eds. *Freedom and Moral Responsibility: General and Jewish Perspectives*, College Park: University of Maryland Press, pp. 85-98.
- Smilansky, Saul (2000). *Free Will and Illusion*, Oxford: Oxford University Press.
- Sommers, Tamler. (2007) "The Objective Attitude," *Philosophical Quarterly* 57, pp. 321-41.
- Spinoza, Baruch (1677/1985). *Ethics*, in *The Collected Works of Spinoza*, ed. and tr. Edwin Curley, Volume 1, Princeton: Princeton University Press.
- Strawson, Galen (1986). *Freedom and Belief*, Oxford: Oxford University Press.
- Strawson, Peter F. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962), pp. 1-25.
- Stump, Eleonore (1996). "Libertarian Freedom and the Principle of Alternative Possibilities," in *Faith, Freedom, and Rationality*, ed. Jeff Jordan and Daniel Howard Snyder, Lanham, MD: Rowman and Littlefield, pp. 73-88.
- Taylor, Richard (1966). *Action and Purpose*, Englewood Cliffs NJ: Prentice-Hall.
- Taylor, Richard (1974). *Metaphysics*, 4th ed., Englewood Cliffs NJ: Prentice-Hall.
- van Inwagen, Peter (1983). *An Essay on Free Will*, Oxford: Oxford University Press.

- Vihvelin, Kadri (1991). "Freedom, Causation, and Counterfactuals," *Philosophical Studies* 64 (1991), pp. 161-84.
- Vihvelin, Kadri (2004). "Free Will Demystified: A Dispositional Account," *Philosophical Topics* 32 (2004), pp. 427-50.
- Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*, Cambridge, Harvard University Press.
- Waller, Bruce (1990). *Freedom without Responsibility*, Philadelphia: Temple University Press.
- Warfield, Ted (2000) "Causal Determinism and Human Freedom are Incompatible," *Philosophical Perspectives* 14 (2000), pp. 167-80.
- Watkins, Eric (2005). *Kant and the Metaphysics of Causality*, Cambridge: Cambridge University Press.
- Watson, Gary (1975). "Free Agency," *Journal of Philosophy* 72, pp. 205-20.
- Watson, Gary (1987). "Responsibility and the Limits of Evil," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman, Cambridge: Cambridge University Press, 1987, pp. 256-286.
- Widerker, David (1995). "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review* 104, pp. 247-61.
- Widerker, David (2000). "Frankfurt's Attack on Alternative Possibilities," *Philosophical Perspectives* 14, pp. 181-201.
- Widerker, David (2006). "Libertarianism and the Philosophical Significance of Frankfurt Scenarios," *Journal of Philosophy* 103, pp. 163-87.
- Wolf, Susan. (1990). *Freedom within Reason*, Oxford: Oxford University Press.
- Wood, Allen (1984). "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, Ithaca NY: Cornell University Press, pp. 73-101.
- Wyma, Keith D. (1997). "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34, pp. 57-70.
- Yaffe, Gideon. *Manifest Activity: Thomas Reid's Theory of Action*, Oxford: Oxford University Press.