

## **Moral Responsibility Without Libertarianism**

Lynne Rudder Baker

After centuries of reflection, philosophers still cannot seem to put to rest the matter of free will. (Compton 2001) The issue of free will remains vital largely because of its connection to moral responsibility. What features are required for moral responsibility? Do we have those features? Questions about moral responsibility are intimately connected to questions about social policy and justice; so, the issue of moral responsibility—of desert, of whether or not anyone is ever really praiseworthy or blameworthy—has practical as well as theoretical significance. Moreover, recent work on free will opens the gate for progress on moral responsibility.<sup>1</sup> I want to try to consolidate the gains and to push forward toward a formulation of adequate conditions for moral responsibility.

There are two familiar main camps: one of compatibilists and the other of incompatibilists. Compatibilists take the conditions for moral responsibility to be compatible with determinism—where determinism is the thesis that the future is completely determined by the laws of nature together with antecedent conditions. Incompatibilists take the conditions for moral responsibility to be incompatible with determinism. Incompatibilists divide into two further groups: those who take us to meet the conditions for moral responsibility (called ‘libertarians’), and those who do not (called ‘hard determinists’).<sup>2</sup> My interest here is mainly with the compatibilists and the libertarians, who agree that we are (at times) morally responsible for what we do, although they differ in what they require for moral responsibility.

The terms ‘compatibilism’ and ‘libertarianism’ standardly refer to positions on the question of free will. However, compatibilist vs. libertarian accounts of free will lead to compatibilist vs. libertarian conditions of moral responsibility. Since, as I said, my concern here is with the conditions for moral responsibility, I’ll use the terms ‘compatibilism’ and ‘libertarianism’ to distinguish two positions on what is required for moral responsibility. As we shall see, what libertarians insist on, and what compatibilists deny, is that moral responsibility for

an action requires that the agent be the source or originator of the action in a way that precludes determinism.

I shall defend compatibilism in two steps: First, I shall argue that libertarianism is false: no one has libertarian freedom. So, if moral responsibility entails libertarianism, then we are never morally responsible for anything that we do. Second, since I do believe that we are morally responsible for certain of our actions, I shall propose nonlibertarian (i.e., compatibilist) conditions for moral responsibility—a Reflective-Endorsement view. I shall add to Harry Frankfurt’s compatibilist conditions to make them sufficient for moral responsibility. Then, I shall defend compatibilism against a recent sustained attack by Derk Pereboom.<sup>3</sup> Finally, I shall comment on the pervasiveness of luck. Along the way, I shall show how compatibilism can accommodate certain libertarian intuitions.

### **What is Libertarianism?**

One of the advances in the free-will debate that I alluded to earlier is the increased detail in which libertarians formulate their positions. Here are some recent formulations of free will by prominent libertarians:

- (i) Free will entails “the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes.” (Kane 1998: 4.)
- (ii) Free will with respect to an action entails that the “action is causally brought about by something that (a) is not itself causally brought about by anything over which she has no control, and (b) is related to her in such a way that, in virtue of its causing her action, she determines which action she performs.” Clarke 1993: 203).
- (iii) Free will with respect to an action entails that the agent’s “own intellect and will are the sole ultimate source or first cause of her act.” (Stump 2001: 126.)
- (iv) Free will with respect to an action X at time t entails that it is within the agent’s power at t both to perform X and to refrain from performing X. (Plantinga 1974: 165-6.)

(v) Free will with respect to an action entails that “no causal laws and antecedent conditions determine either that he will perform the action, or that he will not.” (Plantinga 1974: 166; cf. van Inwagen 1975: 188.)

(vi) Free will with respect to an action entails that the action “is not causally determined by anything outside the agent.” (Stump 2001: 125.)

Whether they regard incompatibility with determinism as a defining condition, or only as a consequence of the nature of free-will, all libertarians agree with (v) and (vi) that an exercise of free will is not causally determined by anything outside the agent. To put it metaphorically, libertarians hold that the causal conditions that produce the action or decision do not have effects that go through the agent (or the agent’s desires) to circumstances outside the agent’s control, but new conditions for the action or decision originate within the agent (or the agent’s desires).

The other formulations suggest two distinct underlying intuitions behind libertarianism. Condition (iv) suggests the Principle of Alternative Possibilities: An agent has free will with respect to an action (and is morally responsible for it) only if the agent could have done otherwise in a nonhypothetical sense. There have been endless discussions over the proper interpretation of ‘could have done otherwise.’ (Berofsky 2002) Since I think that everything that *can* be said about ‘could have done otherwise’ has been said, and also that the Principle of Alternative Possibilities has been decisively refuted,<sup>4</sup> I shall put aside the controversy over the Principle of Alternative Possibilities and focus on the other libertarian intuition—one which, if correct, supersedes the Principle of Alternative Possibilities anyway.<sup>5</sup>

(i), (ii), and (iii) suggest a different libertarian intuition: An agent has free will with respect to an action (and is morally responsible for it) only if the agent has power or control of her actions in an especially weighty way. (Randolph Clarke calls the condition in (ii) the “condition of production.” Clarke 1993: 203.) What kind of control is at issue for the libertarian? In ordinary English, we say that the driver had control of the car (Fischer 1994: 132ff), or that the professor had the power to give all As and the power to refrain from giving all As. These ordinary examples of power or control, however, are clearly compatible with

determinism. The kind of power or control that a libertarian demands is much stronger than our ordinary concept yields. The requisite power or control is originative. The libertarian free agent has power or control over the ultimate sources of her action.

Free will, according to libertarians, requires that the choice or action not have its origin in anything beyond the agent's control. Otherwise, according to prominent libertarian Robert Kane, "the action, or the agent's will to perform it, would have its source in something the agent played no role in producing. Then the arche [sufficient ground or cause or explanation] of the action, or of the agent's will to perform it, would not be 'in the agent,' but in something else." (Kane 1998: 35.)

"[T]o will freely, in this traditional sense," Kane says, "is to be the ultimate creator (prime mover, so to speak) of your own purposes." (Kane 1998: 4.) Roderick Chisholm, a proponent of agent-causation, put it this way:

If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing—or no one—causes us to cause those events to happen.<sup>6</sup>

One of the recent advances in the free-will debate has been disentanglement of the Principle of Alternative Possibilities from the condition of originative control endorsed by Kane, Chisholm, Clarke and others. Although compatibilists have led the attack on the Principle of Alternative Possibilities, some libertarians have pointed out that one can reject the Principle of Alternative Possibilities, without rejecting libertarianism. (Hunt 2000; Kane 1998, Ch. 3; Stump 1999; Pereboom 2001, Ch. 1.) The deeper libertarian intuition than the one codified in the Principle of Alternative Possibilities is that free agents are ultimate originators of their willings and actions. The attacks, led by Frankfurt, on the Principle of Alternative Possibilities do not threaten this deeper libertarian conception of freedom and moral responsibility at all.<sup>7</sup>

The deeper libertarian conception of free will and moral responsibility is thus not a matter of alternative possibilities, nor is it simply a matter of indeterminism. The mere

occurrence of an undetermined event in the causal history of an action does not make the action free by anyone's lights. (We need to know what 'the power to do A or to refrain from doing A' consists in if it is not the ordinary power—say, to wear boots or to refrain from wearing boots.) The fact that the causal history of an action contains an event governed by a statistical law rather than a universal law cannot render an agent morally responsible for the action. Nor, on a non-nomological conception of causation, would the fact that an action or decision was caused by a random event render an agent morally responsible for it. An indeterministically caused decision is no more within the agent's control than is a deterministically caused decision. What is needed—as recent work on libertarianism has shown—is the notion of ultimate origination or ultimate control.

But what is ultimate origination? Consider Kane again, as he speaks of “the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes:”

[W]hen we trace the causal or explanatory chains of action back to their sources in the purposes of free agents, these causal chains must come to an end or terminate in the willings (choices, decisions, or efforts) of the agents, which cause or bring about their purposes. If these willings were in turn caused by something else, so that the explanatory chains could be traced back further to heredity or environment, to God, or fate, then the ultimacy would not lie with the agents but with something else. (Kane 1998: 4.)

This suggests that an agent is the ultimate originator of a choice or action if and only if the sufficient condition for the choice or action includes something over which the agent has complete control. Ultimate originators have ultimate control. Of course, agents do not have ultimate control over all the necessary conditions for their actions. However, if there is a sufficient condition for the choice or action over which the agent lacks ultimate control, the agent is not the ultimate originator of it.

Perhaps someone will object that libertarianism does not require that agents of free actions be ultimate originators. I would reply: Not only have I cited a number of libertarians who are explicitly committed to the idea of ultimate origination, but also without ultimate origination, the libertarian would have no better grounding for moral responsibility than does the

compatibilist. Beliefs, desires, character, upbringing, genetic make-up and so on are outside the agent's control. If you remove the requirement of ultimate origination and suppose that these causal factors are sufficient for an agent's free action or choice, your view is tantamount to compatibilism since these causal factors are beyond the agent's control. So, ultimate control is required for libertarianism.

Before turning to arguments against libertarianism, let me give an example to suggest that our moral practices are at odds with libertarianism: In the 19<sup>th</sup> century, the institution of slavery set the prevailing ethos of the American South. Children were typically brought up strictly in an environment dominated by race. Many 19<sup>th</sup>-century U.S. Southerners believed that slavery is sanctioned by God (part of the "orders of creation"), and by the U.S. Constitution and by common sense.<sup>8</sup> In 1863, a cavalry lieutenant from Mississippi reaffirmed his belief that "this country without slave labor would be wholly worthless....We can only live & exist by that species of labor: hence I am willing to continue the fight to the last." (McPherson 2001: 32.) Given his beliefs and desires—and the fact that there are no competing ones—it would have been irrational for him to have stopped fighting, whether he had libertarian freedom or not and whether he had ultimate control over his decisions or not. Since I assume that the lieutenant was acting rationally on this occasion, even if he could have done otherwise, he *would not* have done otherwise—not without a change in antecedent conditions.<sup>9</sup>

But our practices hold morally responsible people who fight for a cause because they decide to without coercion—with no inquiry into whether they have had ultimate control over their decisions. (Indeed, it is difficult to see how such an inquiry could even proceed.) Ultimate control is simply irrelevant to our moral practices. The lieutenant indicates that his beliefs and desires are sufficient for his intention to "continue to fight to the last," and our ordinary moral practices take them to be sufficient. Otherwise, before holding someone morally responsible for an action, we should inquire into its history to see whether there was some "input" over which the agent had ultimate control. In any case, a libertarian must either deny that the lieutenant was morally responsible for fighting, or else reject the (plausible) description of the case, a description that fully accounts for his action in terms of attitudes and values. There would be a high cost for our moral practices either way.

Let us now turn from pragmatic considerations to direct arguments against libertarianism. In the next section, I shall argue that no human person is ultimate originator of any action, decision or choice.<sup>10</sup>

### A Critique of Libertarianism

There are two main kinds of libertarian theories: Agent-causal theories and event-causal theories.<sup>11</sup> According to agent-causal theories, the causation of “free choices or actions cannot be explained as the causation of events or occurrence by other events or occurrences.” (Kane 1989: 118.) An agent-causal free action “is such that its occurring rather than not here and now, or vice versa, has as its ultimate or final explanation the fact that it is caused by the agent here and now.” (Kane 1989: 120. Emphasis his.) According to event-causal libertarian theories, all causation is causation by events. Event-causal libertarian theories hold that for an agent to cause her free action or choice is for there to be an event involving the agent that causes the action or choice. But both kinds of libertarian theories agree that a free action or choice has its *final or ultimate* explanation in the agent herself: The agent produced it by, e.g., making an effort of will (event-causal theories) or the agent just caused it *simpliciter* (agent-causal theories).<sup>12</sup> There is no further explanation to be had.

Event-causal libertarians like Kane take indeterminism to make room for our being ultimate originators. Agents in an undetermined world have “the power to make choices which can only and finally be explained in terms of their own wills.” (Kane 1989: 129; emphasis his.) Kane takes efforts of will to be indeterministic processes that terminate intentionally in a choice that is undetermined. These indeterministic efforts of will (Should I or shouldn't I?) are correlated with indeterministic macroprocesses in the brain, which in turn result from amplification of microindeterminacies. (Kane 1989: 129) Kane says that when “the indeterminate effort becomes determinate choice, the agent will make one set of reasons or the other prevail then and there by deciding.” (Kane 1989: 134; emphasis his.) But this raises a

question of intelligibility: How can an undetermined event that terminates an indeterministic physical process be produced by an agent's deciding? It is wholly obscure how an agent can intentionally bring an indeterministic brain process to a conclusion by making a choice.<sup>13</sup> Since Kane's view is the most detailed formulation of event-causal libertarianism that I know of, I do not believe that event-causal libertarianism will allow an agent to be the ultimate originator of an action.

In general, we have no more control over an event produced by an indeterministic process than we do over an event produced by a deterministic process. Part of the motivation for libertarian theories is that libertarians take determinism to threaten free will as they see it. The present point, not original with me, is that if determinism threatens free will, then so does indeterminism. (See, e.g., Strawson 2000; Double 1991; van Inwagen 1975.)

Agent-causal theories also take agents to be ultimate originators. Clarke echoes Chisholm's conception of an agent as an "uncaused cause." (Clarke 1993: 201.) Before giving a general argument against the whole libertarian conception of free agents as ultimate originators, let me note that agent-causal theories also founder on a different shoal. According to agent-causal theories, there is a fundamental causal power that allows agents to be the source (i.e., to cause) their free decisions and actions. The exercise of this agent-causal power cannot be reduced to causation between events. The agent, a substance, is the cause of an event—not in virtue of any event involving the agent, but simply by exercise of agent-causal power. In this way, the agent can initiate new causal chains—causal chains with a first event caused directly by the agent. The problems with agent-causal theories are well-known: The theories are obscure; they simply postulate a power that meets libertarian conditions for free will without making plausible how we could have such an amazing power. Although I am sympathetic to the claim that there are emergent properties, appeal to emergent properties in this context does not explain (to me) how there could be libertarian agent-causation. (See O'Connor 1995; O'Connor 2000.) The other problem is that agent-causation seems at odds with the dominant worldview that is broadly "naturalistic."

Although there are many versions of “naturalism” with which I have little sympathy, I differ from some traditional metaphysicians who want to insulate metaphysical theories from developments of science.<sup>14</sup> There is a lot of room between Chisholm and Quine. Without going so far as to regard science as the arbiter of reality, the successes of the sciences give us good reason to take the sciences to be the source of knowledge of physical mechanisms. Although the brain, a physical organ, still seems largely a mystery (especially in relevant areas of, say, complex motivation), enough is known about the brain in broad outline to make agent-causation wholly implausible, or seemingly magical.<sup>15</sup> If the ‘agent-causation’ hypothesis were true, then there should be neural gaps—brain events with no causal (event-)antecedents. In order to be useful to proponents of agent causation, such neural gaps must be different from the ordinary quantum indeterminacy found in inanimate objects. But despite burgeoning neuroscientific research on free will, there is no evidence of relevant neural gaps.<sup>16</sup> Therefore, we should reject agent-causal libertarian theories.<sup>17</sup>

The libertarian holds that if an action is free, it is not explainable by the sum of the kinds of causal conditions that psychologists and social scientists appeal to—such as how an agent sees her situation, what she takes her options to be, her other beliefs, her desires, her character, her experience, her genetic inheritance, her environment. Taken together, these cannot be sufficient for a libertarian free action. Libertarian free action awaits exercise of an ability that no natural being has: the ability to rise above the complex mix of causes (heredity, environment, beliefs, desires, etc.) and interject an unexplainable X factor, over which the agent has ultimate control, and which renders theretofore-insufficient causes sufficient for the choice or action. An ultimate originator seems not to be a part of the natural order.

In short: Ultimate origination is an illusion. No one has it. But without ultimate origination or ultimate control, libertarianism would be on a par with compatibilism vis-à-vis moral responsibility. So, we are justified in rejecting libertarianism.

### **Compatibilist Conditions for Moral Responsibility**

Much of the interest in the question of free will, as I mentioned at the outset, stems from concern about moral responsibility. Some philosophers believe that moral responsibility requires libertarianism. Such philosophers should conclude that no one is ever morally responsible for anything.<sup>18</sup> But logically speaking, such a dire conclusion is not forced: What follows from the argument against libertarianism is only this conditional:

If libertarianism is necessary for moral responsibility, then we are not morally responsible for what we do.

As we are all aware, one philosopher's modus ponens is another philosopher's modus tollens. And such is the case here. From the conditional, one may conclude, via modus ponens, that we are not morally responsible for anything that we do. One may equally well conclude, via modus tollens, that libertarianism is not necessary for moral responsibility. I shall argue for moral responsibility without libertarianism. To do this, I shall set out and defend nonlibertarian conditions for moral responsibility in terms of reflective endorsement. Incompatibilists typically charge that compatibilist conditions for moral responsibility are only necessary, not sufficient. I'll try to meet this charge by supplying sufficient conditions for moral responsibility that are compatible both with determinism and with indeterminism.<sup>19</sup>

Harry Frankfurt, who will provide my starting point, has elaborated on the conditions for moral responsibility. He offers a hierarchical view of the will: We not only have first-order desires; we also have second-order desires. (Frankfurt, 1971) We not only want to do certain things; we also want to want certain things, to be moved by certain desires. A person's will is the desire which actually motivates him: To will X is to have an effective desire to choose or do X, a desire that moves one to choose or to do X. If one has the will that one wants to have (if, that is, one is moved by the desire that one wants to move one), then on Frankfurt's view, one has (compatibilist) free will.<sup>20</sup>

One important feature of Frankfurt's view of the hierarchical will that has gone largely unremarked is that it requires that the agent have a first-person perspective. A person must be able to conceive of her desires *as her own*—from the first-person—if she is to desire to have a certain desire. If Sally wants to be moved by a certain desire, then she wants that she herself

(considered from the first-person) be moved by that desire. To want that someone who is in fact herself be moved by that desire is not enough. If I can make the distinction between wanting that I (myself) be moved by a certain desire and wanting that S (where ‘S’ is a third-person name for me) be moved by that desire, then I have a first-person perspective.<sup>21</sup> As we shall see, only beings with first-person perspectives can be morally responsible.

On my view, a first-person perspective is the defining characteristic of persons.<sup>22</sup> A first-person perspective is the ability to entertain a certain kind of thought about oneself—not just a thought about someone who happens to be oneself, but about oneself as oneself: the person *per se*, without recourse to any name or description. Oedipus’s realization that he himself was the killer of Laius, as opposed to realizing that a man at the crossroads was the killer of Laius, required a first-person perspective. As the example of Oedipus indicates, first-person perspective is irreducible to a third-person perspective. A first-person perspective is what is distinctive about persons.<sup>23</sup>

The importance of the first-person perspective for moral responsibility is that it gives us limited control over our desires. A first-person perspective makes it possible for us, unlike the other animals, to discover what goals we have, to evaluate them and to try to change them. Animals without first-person perspectives do not have this control: they act on their desires, but they cannot set about changing them. But persons can even interfere with biological goals of survival and reproduction. To a significant extent, a person can know what desires she has: she approves of some and is repelled by others. We have partial control, albeit limited by the kinds of people we already are, over our desires. Our partial control is manifest in our (sometimes successful) efforts to change. Yet, such effort—which is limited by our heredity, environment, experience, even by our insight and imagination—is itself a product of factors beyond our ultimate control.

This partial control over our desires has the consequence that, even if determinism is true, we are not mere conduits for causes beyond our control. Because of our first-person perspectives, we have a hand in what causes our actions: we can modify our first-order desires that produce our intentions and in this way to help shape the causes of what we do. We can

decide to try to be one kind of person rather than another, to be generous rather than stingy, say. For incompatibilists—both libertarians and hard determinists—control is all or nothing: either we have ultimate control or we are mere puppets. Attention to the first-person perspective should make us reject this dichotomy: we are neither puppets nor ultimate originators. We can try to become the people that we want to be—even though our wanting to become a certain kind of person, along with the trying itself, is caused by factors beyond our control.

The first-person perspective cannot be acquired by neural manipulation—any more than a disposition to be honest or the ability to read French or any other intentional disposition can be acquired by neural manipulation. I am supposing that a person at *t* (when he is not actually reading French) may have the ability to read French without there being anything in his brain at *t* that makes it the case that a person can read French. I doubt that putting the brain in a given state is ever sufficient for having an intentional disposition. Someone can have a first-person perspective only if he has consciousness and has had many kinds of intentional states. Moreover, distinct first-person perspectives may be qualitatively similar.

Frankfurt's conditions for moral responsibility clearly require a first-person perspective: "Suppose that a person has done what he wanted to do, that he did it because he wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted." (Frankfurt 1971): 19.) Then, if one "identifies" (Frankfurt's word) oneself with a first-order desire that motivates one to do *X*, then, at least *prima facie*, one is responsible for doing *X*. Every piece of this account requires a first-person perspective. Using a '\*' to mark our attribution to the agent of a the first-person perspective: *S*'s doing what he\* wanted to do, *S*'s doing it because he\* wanted to, *S*'s having the will that he\* wanted to have, *S*'s identifying himself\* (not equivalent to *S*'s identifying *S*) with a first-order desire that motivated him—all these are manifestations of a first-person perspective. So, Frankfurt's conditions for moral responsibility require that the agent have a first-person perspective.

Suppose that in addition to satisfying Frankfurt's conditions, the agent endorsed her desires, knowing that they resulted from things that happened before she was born; suppose that

she identified herself with the factors beyond her control that contributed to her desires and beliefs. More precisely, the condition that I want to add to Frankfurt's conditions is this:

- (a) S has the capacity to consider the sources of her\* desires.
- (b) If (i) S had known that her\* wanting to will X had causal antecedents that traced back to factors beyond her\* control, and
  - (ii) S had known of the causal antecedents that traced back to factors beyond her\* control that they were in the causal history of her\* wanting to will X and that they were beyond her\* control, then:  
S still would have: willed X and wanted to will X and willed X because she\* wanted to will X.

I shall abbreviate this cumbersome extra two-fold condition by this: "S would still have wanted to will X even if she\* had known the provenance of her wanting to will X." Now we have compatibilist conditions for free will, which, I submit, become sufficient conditions for moral responsibility. I'll call this view the Reflective-Endorsement view:

- (RE) A person S is morally responsible for a choice or action X if
  - X occurs and:
    - (i) S wills X,
    - (ii) S wants that she\* will X [i.e., S wants to will X],
    - (iii) S wills X because she\* wants to will X, and
    - (iv) S would still have wanted to will X even if she had known the provenance of her\* wanting to will X.

Conditions (ii) – (iv) each require that S have a first-person perspective. Moreover, these conditions are compatible with the truth of determinism. The causal histories of S's desires and willings may be traceable to factors outside of S's control. Even if S could have done otherwise, she would not have done otherwise. Since "the will that moved her when she acted was her will because she wanted it to be, she cannot claim that her will was forced upon her or that she was a passive bystander to its constitution."<sup>24</sup>

As an illustration of the Reflective-Endorsement view, consider the real-life case of Bobby Frank Cherry, who was recently convicted in the bombing that killed four black Sunday-School girls in a church in Birmingham in 1963. Suppose that Cherry (i) willed to participate in the bombing. As a convinced white supremacist (who apparently bragged of his participation), he (ii) wanted to will to participate, and he (iii) willed to participate because he wanted to. He (iv) would still be proud of his participation, and would participate again, even though he knew that his wanting to will to participate in the bombing had been caused by his racist upbringing. (“Damn right,” he might have said, “and I’m bringing up my boys the same way.”) It seems to me obvious that he was morally responsible for his participation in the bombing. He was moved by the desire (to bomb the black church) that he wanted to be moved by. Cherry satisfies (i) – (iv) of (RE), and (I believe that) we properly find him morally responsible.

A basic conviction—shared by both libertarian and compatibilist—is that moral responsibility concerns the agent in a deep way. An agent is morally responsible for an action if he *endorses* the beliefs and desires on which he acts: When he affirms them as his own (and makes no factual errors about the circumstances, etc.), he is morally responsible for acting on them. Whereas the libertarian is concerned that a morally responsible agent have (impossible) control over factors that contribute to what he wills, the compatibilist is concerned that a morally responsible agent endorse and align himself\* with what he wills. Doing what we want to do, with reflective endorsement, is all the control that we have, and that’s enough.

If I can say, “These desires reflect who I am, and this is the kind of person that I want to be,” then (surely!) I am morally responsible for acting on those desires—whether determinism is true or not. I have (nonultimate) control of what I do if, in the absence of psychopathology like kleptomania, I deliberate and make an uncoerced decision to do it.<sup>25</sup> Unlike the libertarian, the compatibilist does not think that external causes of decisions or actions per se threaten moral responsibility; moral responsibility is precluded only by certain kinds of external causes—those that by-pass the agent’s own psychological contribution involving his first-person perspective. By focusing on the first-person perspective, a compatibilist can agree with the libertarian to this extent: Being morally responsible has to do with the agent in a deep way, not just with an

external series of causes. Reflective Endorsement offers a via media between a (physically dubious) libertarianism, and a (morally hopeless) hard determinism.

### **A Challenge to Compatibilism**

Now I want to turn to a challenge to the Reflective-Endorsement view. Derk Pereboom does not believe that compatibilists can give sufficient conditions for moral responsibility. He uses a “combined counterexample and generalization strategy” to undermine various compatibilist accounts. Assuming for the sake of argument that determinism is true, Pereboom offers a graduated series of four cases of Professor Plum’s killing Ms. White for some personal gain, and then asks where to draw the line between Plum’s not being morally responsible and Plum’s being morally responsible for killing Ms. White. (Professor Plum and Ms. White are two characters in a board game called ‘Clue.’) Here are skeletons of the four cases:

Case 1: Plum is created by neuroscientists, who directly manipulate his reasoning via radio-like technology, “directly producing his every state from moment to moment....His effective first-order desire to kill Ms. White conforms to his second-order desires.” [Pereboom 2002: 112-3.]

Case 2: Plum is created by neuroscientists and programmed to weigh reasons for action. The programming causally determines him to “to possess the set of first- and second-order desires that results in his killing Ms. White.” [Pereboom 2002: 114.]

Case 3: Plum is an ordinary human being, “except that he was determined by the rigorous training practices of his home and community,” and these practices “determined his character.” [Pereboom 2002: 114.] Again, he has the first- and second-order desires that lead to his killing Ms. White.

Case 4: Plum is an ordinary human being, raised in normal circumstances. He “exhibits the specified organization of first- and second-order desires.” [Pereboom 2002: 115.]

Pereboom’s claim is that Plum is not morally responsible in Case 1 and there are no morally significant differences between any two adjacent cases. He argues that if Plum is not

morally responsible in Case 1, then he is not in Case 2; and if he is not morally responsible in Case 2, then he is not in Case 3; and if he is not morally responsible in Case 3, then he is not in Case 4. Pereboom concludes that since Case 4 is just the normal case, the verdicts on the sequence of counterexamples generalize: The compatibilist conditions are never sufficient for moral responsibility. Since Pereboom also believes that libertarianism is false, he concludes that there is no moral responsibility, period.

I want to show that the Reflective-Endorsement view (RE) can stand up to Pereboom's strategy. It is inevitable that my treatment of these Cases relies, to a greater extent than I would like, on intuition—just as Pereboom's treatment does. What is at issue is whether the conditions of (RE) are sufficient for moral responsibility. Is everyone who satisfies these conditions morally responsible? Whether or not failure to satisfy them relieves one of moral responsibility is a different question, not at issue here.<sup>26</sup> So, to test the sufficiency of (RE), we need to see whether in every case in which the conditions of (RE) are satisfied, the agent is intuitively morally responsible. If so, (RE) withstands Pereboom's argument; if not, (RE) fails.

The complication of Pereboom's strategy is that it asks us where, in the series of four cases, Plum becomes morally responsible for killing Ms. White. This question does not have a straightforward answer, because Cases 3 and 4 are open to multiple interpretations.<sup>27</sup> I shall divide Cases 3 and 4 into "a" and "b" interpretations. The interesting interpretations are 3b and 4b. I shall just quickly mention 3a and 4a in order to put them aside. My tack will be to show that Plum never—on any interpretation of any of the four cases—satisfies all the conditions of (RE) without being morally responsible.

Case 3a: Interpret Case 3 so that Plum does not know the source of his desire to kill Ms. White, and has been rebelling against his rigorous upbringing. If he were to realize where his murderous desire came from, he would no longer want to have that desire as his will. In that event, (iv) of (RE) would remain unsatisfied.<sup>28</sup> An example in which (iv) of (RE) is unsatisfied does not impugn the sufficiency of the conditions of (RE). So put Case 3a aside.

Case 4a: Suppose that Plum were to discover the sources of his desire to kill Ms. White and the related second-order desires (perhaps through psychoanalysis), and that that knowledge

disgusted him. He consequently repudiated those desires, in which case condition (iv) of (RE) would not be satisfied. (RE) would not count him as morally responsible in Case 4a. Hence, Case 4a is also irrelevant to the sufficiency of the conditions of (RE) for moral responsibility, and so put it aside too. The four cases for further response are Case 1, Case 2, Case 3b, and Case 4b.

Response to Case 1: Everyone agrees that Plum in Case 1 is not morally responsible. At best, the neuroscientists are simulating “identification at the level of second-order desire with his will,” and simulating reasoning in Plum. But since there is direct manipulation moment by moment, Plum is not reasoning, nor is he identifying any second-order desires with his will. The reasoning is the reasoning of the manipulator, not of Plum: If the manipulate went home for the night, the appearance of reasoning in Plum would disappear. Moreover, in order to have second-order desires at all, Plum must have a first-person perspective, and he must have all the empirical concepts that he needs to decide to kill Ms. White. There is good reason to deny that it is physically possible to create a being with a first-person perspective simply by directly manipulating his reasoning at every step of the way.<sup>29</sup> (Indeed, it is difficult to see how a directly-manipulated being would even have empirical concepts—any more than a brain in a vat would have empirical concepts.) So, (ii) and (iii) of (RE) are unsatisfied in Case 1. Plum is no more reasoning than the paralyzed patient is moving her arm when the arm is being lifted by a physical-therapy machine. Nor does Plum have the capacity to consider the sources of his desires and thus does not satisfy (iv) either. Intuitively, Plum in Case 1 is not morally responsible, and the Reflective-Endorsement view agrees.

Response to Case 2: Assuming that Plum is just behaving in accordance with a program written by the neuroscientists and that the program is being executed as written (i.e., there is no “learning;” the program does not “evolve” on its own), Case 2 is morally similar to Case 1. Plum still does not satisfy (iii) of (RE). He does not will to kill Ms. White because he wants to will it, but rather because he is executing a nonevolving program. A being S with a first-person perspective can distinguish between wanting that S will X and wanting that he himself will X. An inflexible algorithm, not even one modifiable by feedback, could not underwrite this distinction.<sup>30</sup> So, Plum in Case 2 does not have the first-person perspective required by (ii) and

(iii)—any more than a simple program that printed out “I am glad that I am a computer” has a first-person perspective.

Case 1 is to Case 2 as a child’s doll that emits crying noises when you lay her down is to another doll whose emission of crying noises is controlled by a timer. Plum is no more an independent agent in Cases 1 and 2 than is either of the dolls. So, again, Plum in case 2 is intuitively not morally responsible, and (RE) agrees.

Response to Case 3b: On this interpretation of Case 3, suppose that Plum is well aware of the effects of his strict upbringing, but he does not even want to repudiate the resulting attitudes. If Plum is aware of the effects of his upbringing, then he has a first-person perspective. In that case, Plum may well satisfy the conditions of (RE): He wills to kill Ms. White because he wants to, and he may even know that his desires have their source in his strict upbringing, and even be proud of them. In that event—in which (i) – (iv) were satisfied—Plum would be morally responsible. Bobby Frank Cherry is a good real-life example of Case 3b.

Here is another real-life example that may be assimilated to Cases of type 3b. A man said to be a recruiter and dispatcher of suicide bombers on the West Bank was described by a reporter like this: “He grew up in Hamas. It was an ideology for him from the time he was 10. I don’t think he ever felt that he was choosing one route over another. I think he saw the path he took as his identity.” [Rubin 2002: 6] On the Reflective-Endorsement view, seeing one’s identity in a certain path is clear evidence that one is morally responsible for taking that path.

Pereboom says that a compatibilist who takes Plum to be morally responsible in Case 3—Case 3b to us; Plum does not satisfy (RE) on Case 3a—must show how Case 3 differs from Case 2. On the Reflective-Endorsement view, there are large salient differences between Case 2 and Case 3b. The most obvious and important difference is that in Case 3, but not in Case 2, Plum has a first-person perspective. In Case 2, Plum does not will to kill Ms. White because he wants to will it, but because a pre-set program is being executed; nor is he able to consider the sources of his desires, as (iv) requires. In Case 3b, not only does Plum will to kill Ms. White because he wants to will it (thus satisfying (iii) of (RE)), but also he is proud of having these desires even though he is aware of where they came from (thus satisfying (iv) of (RE)). So, the break in the

series between Plum's not being morally responsible and Plum's being morally responsible occurs between Case 2 and Case 3b. The intuitive support for the difference in moral responsibility between Case 2 and Case 3b is mirrored by the Reflective-Endorsement view.

Again, the Reflective-Endorsement view can capture intuitions that libertarians and compatibilists both have concerning the dignity of the person. Human subjects are morally responsible when they affirm the beliefs and desires that move them *as their own*—even if those beliefs and desires have causes beyond their control. The first-person perspective makes the agent the locus of moral responsibility.

Case 4b: If the knowledge of the sources of his desire to kill Ms. White and the related second-order desires left them all intact, then—as in Case 3b—Plum would be morally responsible for killing Ms. White. And, as before, Plum would satisfy all the conditions of Reflective Endorsement. There is no need to find a difference between Case 3b and Case 4b since Plum satisfies (RE) on both and is morally responsible in both.

To sum up: All the cases in which Plum satisfies (RE) are cases in which he is intuitively morally responsible. In Cases 1 and 2, Plum does not satisfy (RE), nor is he morally responsible for killing Ms. White. In Cases 3b and 4b, Plum satisfies (RE) and is morally responsible. (Cases 3a and 4a are irrelevant to whether or not (RE) gives sufficient conditions for moral responsibility (as are Cases 1 and 2).) The first-person perspective provides a principled difference—tracked by (RE)—between Case 2 and Case 3b.

So, I do not think that Pereboom's counterexample-and-generalization strategy is successful against the Reflective-Endorsement view. Indeed, if we had encountered the four cases in reverse order—beginning with Case 4, and ending with Case 1—I suspect that we would immediately judge that Plum was morally responsible for killing Ms. White in Cases 4 and 3, and not in Cases 2 and 1.

This discussion also indicates how a proponent of Reflective-Endorsement can respond to anti-compatibilist arguments invoking the infamous mad neuroscientist. The skeleton argument against compatibilism is this: A mad neuroscientist manipulates an agent's brain, so that the

agent wills to do X (or whatever would satisfy the compatibilist conditions on offer). Then, when the agent does X simply as a result of the neuroscientist's manipulations, he satisfies the compatibilist conditions for moral responsibility; but he is not really morally responsible for doing X. So, it is claimed, compatibilism is false.

The Reflective-Endorsement view does not succumb to this argument. In the first place, in order to be a candidate for being morally responsible for doing something, a being must have a first-person perspective; and a neuroscientist cannot provide a being with a first-person perspective merely by manipulating his brain (any more than he could provide a being with an ability to speak French by manipulating his brain). In the second place, if the neuroscientist works his will on a being who already has a first-person perspective, then that person either will or will not satisfy condition (iv) of (RE). On the one hand, if the agent whose brain is manipulated by the mad neuroscientist does not satisfy (iv)—if knowing of the manipulation, he would be appalled by his wanting to will X—then the Reflective-Endorsement view does not hold him morally responsible. On the other hand, if the agent does satisfy (iv)—if knowing of the manipulation, he would still want to will X—then the Reflective-Endorsement view holds him morally responsible, as indeed it should.

An incompatibilist may press on: What if satisfaction of condition (iv)—an agent's being being such that he would still want to will X even if he knew the provenance of his wanting to will X—were itself produced by the mad neuroscientist? In this case, the mad neuroscientist would make the agent still endorse his willing X even if the agent knew that the endorsement and the willing were caused by a mad neuroscientist. This possibility prompts me to interpret 'knowing the provenance of the agent's endorsement of his willing X' as including a completeness clause: "There is no further knowledge of the circumstances of the agent's endorsement of his willing X that would lead the agent to repudiate his endorsement of his willing X." Now the agent who would have repudiated his endorsement of his willing X if the mad scientist had not prevented his repudiation does not satisfy (iv) after all, and hence is not deemed morally responsible by (RE). However, an agent who does satisfy (i)-(iv), where (iv) is interpreted as including this completeness clause, would not have repudiated his endorsement of

willing X even if the mad neuroscientist not had prevented his repudiation. Such an agent is morally responsible for X.

So, an agent under the control of a mad scientist either satisfies all the conditions of Reflective-Endorsement or not. In neither case do we have a situation in which an agent satisfies (RE) without being morally responsible.<sup>31</sup> So, I believe that the Reflective-Endorsement view escapes the threat from the mad neuroscientist.

### **Why Moral Responsibility Matters**

Compatibilists are much concerned with moral responsibility.<sup>32</sup> But why? Why not join the hard determinists and abandon moral responsibility altogether? The answer, I believe, is that we need recourse to moral responsibility in order to make sense of moral phenomena. Here's an illustration:

In January 2001, two small-town teenaged boys wantonly and brutally murdered two well-respected Dartmouth University professors. The motive was unclear. There was some talk of robbery; there were surmises about the thrill of killing. The boys did not personally know the victims. In court, the older of the two, who by all accounts was the leader of the operation, showed no remorse at all. He sat stone-faced and sullen. The slightly younger boy wept and managed to address the children of the victims with an apology. It is obvious to me that the boy who expressed regret and repentance for his heinous deed is morally superior to the one who remained cold and remorseless. But if no one were ever actually morally responsible, the one who accepted moral responsibility would simply be a dupe who was just deluding himself.

Of course, some may think that it is useful to inculcate in people a difference between feeling bad that something happened and feeling bad for having done it. Machiavellians (but not I) may want to promote belief in moral responsibility as a socially useful fiction even if they think there is none. And Pereboom argues that abandoning moral responsibility leaves much of morality intact, and he may be right. But it is difficult to see how, from a first-person perspective, one can pursue a moral life without appeal to moral responsibility. It is possible that such phenomena as believing that one is morally responsible are always mistaken, but that seems

to me to be utterly implausible. In short, the reason that moral responsibility matters is that it is needed to make sense of ordinary and ubiquitous moral phenomena.

### **The Pervasiveness of Luck**

Many things beyond our control affect us for good or ill, regardless of what we deserve. These things may be called ‘luck’ (which suggests randomness and purposelessness) or ‘fate’ or ‘Providence’ (which suggests inevitability and purposefulness).<sup>33</sup> I want to hold in abeyance questions about purpose or meaning, and just focus on the bare phenomenon of pervasive arbitrariness: goods and ills—including those that produce our characters and circumstances—are typically not apportioned according to desert. I’ll use words like ‘luck’ and ‘arbitrariness’ for this phenomenon (considered apart from any explanation of it).

Does the pervasiveness of luck preclude moral responsibility? I fully agree with hard determinists about the pervasiveness of arbitrariness—if the sun had not been blazing on Meursault in Camus’s *L’Étranger* or if he had been in a different mood, or if the Arab on the beach that day had stayed home, Meursault would not have murdered the Arab. Yes, I agree that luck is everywhere. But life—with arbitrary differences in natural endowment and environmental advantages—is unfair, from anyone’s point of view. Luck “goes all the way through....[E]verything ends up being ultimately a matter of luck,” insists Saul Smilansky, a hard determinist. (Smilanski 2003: 275.) Agreed. Amazingly, Smilansky writes as if this were news to compatibilists. He thinks that compatibilists fail to acknowledge the pervasiveness of luck, and that that failure makes compatibilism—though necessary for treating people with dignity—shallow. But compatibilists are as likely as anyone else to be aware of the Lucky Gene Club, whose members are those born without crippling defects into loving and financially secure families. It does not take much empathy or insight to realize that even your ability to put to good use your lucky genes is itself a matter of luck. Star athletes who are not even especially reflective routinely refer to their talent as a gift and their opportunities for proper training as good fortune. Arbitrariness is not hidden from the view of compatibilists or of anyone else; it smacks you in the face.<sup>34</sup>

Acknowledgment of luck and arbitrariness should not lead us to incompatibilism. We can (and should) try to ameliorate gross arbitrary disparities, but we cannot eliminate them altogether; nor, in the face of arbitrariness, should we give up morality.<sup>35</sup> Morality requires that there be a morally significant difference (which, incidentally, entails a first-person perspective) between “I’m sorry that I did it” and “I’m sorry that it happened.” We need to invoke moral responsibility to make sense of such a difference. So, the pervasiveness of luck should not induce us to abandon moral responsibility. Indeed, the view that moral responsibility requires sovereignty over luck itself is a libertarian notion that should not survive the refutation of libertarianism.

Rather, the ubiquity of luck should lead us—compatibilists and incompatibilists alike—to temper judgments of moral responsibility with pity (e.g., pity for the child molester who was himself molested as a child). But luck should not drive us away from moral judgments altogether.<sup>36</sup> Recognizing the unfairness of luck, it is a mistake to infer that it is “unfair to blame a person for something not ultimately under her control.”<sup>37</sup> The unfairness of luck does not transfer to unfairness of everything that has luck somewhere in the causal background. To give up moral responsibility in the face of the defeat of libertarianism would be to deny that we are moral beings. But it is as indubitable that we are moral beings as it is that nothing is under our ultimate control.

Religious thinkers have seen clearly the two-fold situation—that nothing is under our ultimate control, and yet we are moral beings.<sup>38</sup> They use the term ‘God’s will’ or ‘particular Providence’ to apply to exactly the same phenomena that Smilansky and I have used ‘luck’ to refer to: the uneven distribution of goods and ills that are beyond our ultimate control. Augustine, Calvin and Jonathan Edwards<sup>39</sup> are all what we now call compatibilists; none denied that we are morally responsible for what we do willingly—even though they insisted that we have no ultimate control. According to such thinkers, God’s will is ultimately arbitrary in the older sense that what happens is what God rules, what He arbitrates, and God does not arbitrate on the basis of any principle (at least not on the basis of any principle knowable by us). So, the effects of God’s will are arbitrary in exactly the way that luck is.

In sum, everyone should recognize that luck (or God's will) is pervasive, regardless of any position on moral responsibility. Proponents of the Reflective-Endorsement view can reject hard determinism while acknowledging ultimate arbitrariness: Reality is tragic; we might wish that it were otherwise, but what's shallow is such wishful thinking. Compatibilism-cum-pity is an appropriate and profound response to the way things really are.<sup>40</sup>

### Conclusion

Libertarianism is untenable, but with the Reflective-Endorsement view, I have offered compatibilist conditions that I believe are sufficient for moral responsibility. The importance of moral responsibility—in our lives as moral beings, in our social practices—makes compatibilism desirable. The availability of sufficient conditions for moral responsibility makes compatibilism feasible. So, I urge acceptance of the Reflective-Endorsement view.<sup>41</sup>

University of Massachusetts at Amherst

### References

Baker, Lynne Rudder (forthcoming a) "Why Christians Should Not Be Libertarians: An Augustinian Challenge," Faith and Philosophy.

\_\_\_\_\_ (forthcoming b) "On Being One's Own Person," in Reasons of One's Own, Maureen Sie, Bert van Den Brink, Marc Slors, eds. Ashgate Publishing.

\_\_\_\_\_ (2003a) "The Difference that Self-Consciousness Makes," On Human Persons, Klaus Petrus, ed. (Frankfurt/London: Ontos-Verlag, 23-40.

\_\_\_\_\_ (2003b) "First-Person Knowledge," The Nature and Limits of Human Understanding: The Gifford Lectures, University of Glasgow, 2001, Anthony J. Sanford, ed., London: T.&T. Clark, 165-184.

- \_\_\_\_\_ (2000) Persons and Bodies: A Constitution View, Cambridge: Cambridge University Press.
- \_\_\_\_\_ (1998) "The First-Person Perspective: A Test for Naturalism," American Philosophical Quarterly 35 (1998): 327-348.
- \_\_\_\_\_ (1981) "Why Computers Can't Act," American Philosophical Quarterly 18 (1981): 157-163.
- Berofsky, Bernard (2002) "Ifs, Cans, and Free Will: The Issues," The Oxford Handbook of Free Will, Robert Kane, ed., New York: Oxford University Press, 181-201.
- Chisholm, Roderick (1966) "Freedom and Action," Freedom and Determinism, Keith Lehrer, ed. (New York: Random House, 11-44.
- Clarke, Randolph (1993) "Toward a Credible Agent-Causal Account of Free Will," Noûs 27: 191-203. Reprinted in Agents, Causes, and Events: Essays on Indeterminism and Free Will, Timothy O'Connor, ed. (Oxford: Oxford University Press, 1995): 201-218. Page references in text to the latter.
- Compton, John J. (2001) "The Persistence of the Problem of Freedom," The Review of Metaphysics 55: 95-115.
- Double, Richard (1991) The Non-Reality of Free Will Oxford: Oxford University Press.
- Ekstrom, Laura Waddell (2002) "Libertarianism and Frankfurt-style Cases," The Oxford Handbook of Free Will, Robert Kane, ed. New York: Oxford University Press, 309-322.
- Fischer, John Martin (1994) The Metaphysics of Free Will Oxford: Blackwell Publishing Company.
- \_\_\_\_\_ (2002) "Frankfurt-type Examples and Semi-Compatibilism," The Oxford Handbook of Free Will, Robert Kane, ed., New York: Oxford University Press, 281-308.

\_\_\_\_\_ and Mark Ravizza (1998) Responsibility and Control: A Theory of Moral Responsibility, Cambridge: Cambridge University Press.

Frankfurt, Harry (1969) "Alternate Possibilities and Moral Responsibility," Journal of Philosophy 66: 829-839.

\_\_\_\_\_ (1971) "Freedom of the Will and the Concept of a Person," Journal of Philosophy 68 5-20.

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

Goetz, Stewart (1997) "Libertarian Choice," *Faith and Philosophy* 14: 195-211.

Haji, Ishtiyaque (1998) Moral Appraisability, New York: Oxford University Press.

Hunt, David (2000) "Moral Responsibility and Unavoidable Action," Philosophical Studies 97: 195-227.

Kane, Robert (1989) "Two Kinds of Incompatibilism," Philosophy and Phenomenological Research 50: 219-254. Reprinted in Agents, Causes, and Events: Essays on Indeterminism and Free Will, Timothy O'Connor, ed. (Oxford: Oxford University Press, 1995): 115-150. Page references in the text are to the latter.

\_\_\_\_\_ (1998) The Significance of Free Will, New York: Oxford University Press.

\_\_\_\_\_, ed. (2002) The Oxford Handbook of Free Will, New York: Oxford University Press.

Libet, Benjamin (2002) "Do We Have Free Will?" in The Oxford Handbook of Free Will, Robert, Kane, ed. (Oxford: Oxford University Press): 551-564

McPherson, James M. (2001) "Southern Comfort," April 12, 2001, New York Review of Books.

O'Connor, Timothy (1995) "Agent Causation," Agents, Causes, and Events: Essays on Indeterminism and Free Will, Timothy O'Connor, ed. Oxford: Oxford University Press, 173-200.

\_\_\_\_\_, (2000) Persons and Causes: The Metaphysics of Free Will, Oxford: Oxford University Press.

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

Perry, John (1979) "The Problem of the Essential Indexical," Noûs 13: 3-22.

Plantinga, Alvin (1974) The Nature of Necessity, Oxford: Clarendon Press.

Rubin, Elizabeth (2002) [No title] New York Times Magazine, June 30, 2002, 6.

Smilanski, Saul (2003) "Compatibilism: The Argument from Shallowness," Philosophical Studies 115: 257-282.

Strawson, Galen (2000) "The Unhelpfulness of Indeterminism," Philosophy and Phenomenological Research 60: 149-156.

Stump, Eleonore (1999) "Alternative Possibilities and Moral Responsibility: The Flicker of Freedom," Journal of Ethics 3: 299-324.

\_\_\_\_\_(2001) "Augustine on Free Will," The Cambridge Companion to Augustine, Eleonore Stump and Norman Kretzmann, eds., Cambridge: Cambridge University Press, 124-147.

Van Inwagen, Peter (1975) "The Incompatibility of Free Will and Determinism," Philosophical Studies 27: 185-199. Reprinted in Free Will, Derk Pereboom, ed., Indianapolis: Hackett Publishing Company, 1997: 184-198. [Page references in the text are to the latter.]

\_\_\_\_\_(1978) "Ability and Responsibility," Philosophical Review 87: 201-224.

\_\_\_\_\_ (2000) "Free Will Remains a Mystery," Philosophical Perspectives 14, Action and Freedom, James E. Tomberlin, ed., Oxford: Blackwell Publishers, 1-19. Reprinted in The Oxford Handbook of Free Will, Robert Kane, ed., New York: Oxford University Press, 158-177.

Walter, Henrik (2001) Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy, Cynthia Klohr, trans., Cambridge MA: MIT/Bradford.

Widerker, David (2002) "Responsibility and Frankfurt-type Examples," The Oxford Handbook of Free Will, Robert Kane, ed. (New York: Oxford University Press, 323-334).

\_\_\_\_\_ (2002) "Farewell to the Direct Argument," Journal of Philosophy 99: 316-324.

Zagzebski, Linda (2000) "Does Libertarian Freedom Require Alternate Possibilities?" Philosophical Perspectives 14: 231-248.

## Notes

<sup>1</sup> Examples of this promising new work include the following: (i) Pereboom 2001; (ii) The detailed work by libertarians (e.g., Kane 1998; O'Connor 1995, Clarke 1993); (iii) The many articles generated in response to Harry Frankfurt's attack on the principle of alternative possibilities. For a sample of the responses, see Part V of Kane, ed. 2002: Fischer 2002, Ekstrom 2002, Widerker 2002. Also see van Inwagen 1978 and Fischer 1994.

<sup>2</sup> 'Hard determinism' is the usual label. But as Pereboom points out, one may think that we have no moral responsibility whether determinism is true or false. So, Pereboom adopts the broader term, 'hard incompatibilist' for his own position. Pereboom 2001: xx—xxv.

<sup>3</sup> See Pereboom 2001, who argues against both libertarianism and compatibilism, and concludes that we are not morally responsible. I believe that his arguments against libertarianism, from which I shall draw, are decisive; but I want to rebut his arguments against compatibilism.

<sup>4</sup> For an important argument—one that spawned an industry of responses—against the Principle of Alternative Possibilities, see Frankfurt 1969. I believe that Pereboom's "new Frankfurt-style scenario" clinches the case against the Principle of Alternative Possibilities. Pereboom 2001: 18-28. Nevertheless, I admit that the Principle of Alternative Possibilities remains controversial—though I do not think that it should.

<sup>5</sup> In any case, David Widerker argues that the notion of avoidability remains central to the incompatibilist's claim that determinism excludes moral responsibility. Widerker 2002.

<sup>6</sup> Chisholm 1966: 23. Emphasis mine. Timothy O'Connor, another agent-causalist mentions that some have used the term 'unmoved movers,' but he prefers 'not wholly moved movers.' (O'Connor 1995: 174). O'Connor wants to allow causal relations between an agent's reason and resulting behavior. On O'Connor's view, agents have a property that can "make possible the direct, purposive bringing about of an effect by the agent who bears it." (p. 177) Emphasis his.

<sup>7</sup> E.g, Linda Zagzebski, an incompatibilist, says: "The presence of alternate possibilities may be a reliable sign of the presence of the agency needed for responsibility, but it is not necessary for it." Zagzebski 2000: 245. Some philosophers agree that Frankfurt-style examples cast doubt on the principle, but they separate blameworthiness from wrongness. E.g., Haji 1998 argues that if determinism is true, one may be blameworthy, but there are no wrong actions.

<sup>8</sup>Mississippi's secession commissioner to Maryland insisted that "slavery was ordained by God and sanctioned by humanity." McPherson, 2001: 30.

<sup>9</sup> Unlike Kant, I do not think that it is always irrational to be immoral.

<sup>10</sup> For theological arguments against the libertarian conception of freedom, see Baker forthcoming a.

<sup>11</sup> These labels are Pereboom's. Kane calls the two types 'Agent Cause Theories' and 'Teleological Intelligibility Theories.' Kane 1995. There is another brand that might be called 'agent-noncausal libertarianism,' according to which "a choice is essentially an uncaused purposeful mental action." On this view, agents have "ultimate and irreducible" mental powers, and "[t]o choose is to exercise the mental power to choose." Goetz 1997: 196. Also see Ginet 1990. Since on these accounts, agents are said to have ultimate mental powers to make uncaused choices, I think that it's fair to say that noncausal libertarian accounts also conceive of free agents as ultimate originators.

<sup>12</sup> Noncausal libertarians would not say that the agent caused the choice, but rather that the choice was the product of the agent's uncaused exercise of the power to choose.

<sup>13</sup> Although this point has been made numerous times, I believe that Pereboom's careful arguments against event-causal libertarianism are decisive. For more detailed arguments against all varieties of libertarianism, see Pereboom 2001.

<sup>14</sup> There are extremely complex methodological issues here that, unfortunately, I cannot pursue in this work.

<sup>15</sup> The reasons to reject agent-causal libertarianism apply equally to agent-noncausal libertarianism.

<sup>16</sup> In light of the *prima facie* implausibility of neural gaps, it would take us off the point to discuss the absence of evidence for them in detail. For some recent results in neuroscience, see Libet 2002 and Walter 2001. For detailed empirical objections to agent-causal libertarianism, see Pereboom 2001: Ch. 3.

<sup>17</sup> Or at least wait for neuroscientific evidence of the relevant gaps. In general, I do not think that philosophy must simply defer to science. However, when a philosophical thesis has empirical implications (as agent-causation does), and the empirical implications are *prima facie* implausible (as relevant neural gaps are), then I think that the absence of scientific evidence in favor of the empirical

implications justifies rejection of the philosophical thesis. Nevertheless, if there were empirical evidence of relevant neural gaps, I would reconsider my opinion.

<sup>18</sup> Derk Pereboom does conclude that no one is morally responsible for anything.

<sup>19</sup> A full account of moral responsibility would provide necessary as well as sufficient conditions. However, since the charge against compatibilism has been that compatibilist conditions are not sufficient for moral responsibility, whether they are necessary or not, the Reflective-Endorsement account is a contribution to the debate.

<sup>20</sup> Frankfurt himself holds that freedom of the will, although compatible with determinism, is not required for moral responsibility. What is at stake in this discussion, however, are sufficient conditions for moral responsibility.

<sup>21</sup> For a general account of the first-person perspective, see Baker 2000. For an application of the account to the issue of moral agency, see Baker forthcoming b.

<sup>22</sup> On my view, the first-person perspective is not just a pragmatic feature of language that can be understood just in terms of indexicals. (Baker 2000). It cannot be assimilated to Perry's important work on the essential indexical. (Perry 1979.)

<sup>23</sup> I have discussed the first-person perspective at length elsewhere. See, in addition to Baker 2000, Baker 2003a, Baker 2003b, and Baker forthcoming b.

<sup>24</sup> Frankfurt 1971: 19. I changed the pronouns to feminine in order to make the sentence grammatically parallel to (i) – (iv).

<sup>25</sup> The conditions of the Reflective Endorsement view may be regarded as ways to accommodate Aristotelian defeating conditions of voluntariness: ignorance, compulsion, reduced capacity and the like. One reason not to focus on Aristotle in the debate over compatibilism is that his view has been claimed by both compatibilists and libertarians.

<sup>26</sup> (RE) gives sufficient conditions for moral responsibility, not an analysis. (RE) does not cover omissions or obligations to one's parents, one's children, one's country that one does not choose to incur. I want to leave open the possibility that in addition to (RE), there are other sets of conditions sufficient for moral responsibility. (Compare: Having a duly authorized person say certain words in a certain setting is sufficient for getting married. But there is another unrelated sufficient conditions: One also gets married by living with a partner of the other sex for seven years. A common-law

marriage is still a marriage.)

<sup>27</sup> Pereboom did not design the four cases with (RE) in mind. But I want to support (RE) by showing that it does not fall to Pereboom's argument.

<sup>28</sup> I take no stand here on whether Plum is morally responsible in this case. Perhaps there are other sufficient conditions for moral responsibility besides (RE). The only point that I am making here is that Case 3a does not impugn (RE).

<sup>29</sup> For arguments to this effect, see Baker 1998.

<sup>30</sup> See Baker 1998. For an earlier attempt to establish this conclusion, see Baker 1981.

<sup>31</sup> Since (RE) is only a sufficient condition, there may be cases in which an agent who does not satisfy (i) – (iv) is nevertheless morally responsible.

<sup>32</sup> Given the arguments against libertarianism, if moral responsibility is indeed not compatible with determinism—as van Inwagen 1975 argued—then we should give up moral responsibility. See also van Inwagen 2002.

<sup>33</sup> I do not intend to disagree here with thinkers like Augustine, Boethius, and Aquinas, who take luck and Providence to be different categories. However, for purposes here, they can be treated together inasmuch as they concern matters beyond an agent's control that are distributed in a way disproportionate to desert.

<sup>34</sup> Smilansky advocates combining compatibilism (at the level of social institutions) with hard determinism (at the ultimate level). I am skeptical of his use of "levels". And the compatibilist can, and should, recognize luck as well as the hard determinist.

<sup>35</sup> Pereboom argues that many of the purposes served by morality can still be served without moral responsibility. Removing moral responsibility seems to me to eviscerate morality.

<sup>36</sup> Nor should luck drive us to the pretense of making moral judgments that we do not really endorse for the sake of a decent society. This sort of pretense, associated with certain political philosophers, I find appalling.

<sup>37</sup> If, as I am confident, this intuition is mistaken, then hard determinism is irrelevant to the fact of ultimate arbitrariness.. Smilansky 2003: 268. (Emphasis his.)

<sup>38</sup> In the Christian tradition, Augustine, Calvin and Jonathan Edwards come to mind in this regard. I suspect that there are comparable figures in the Jewish and Muslim traditions as well.

<sup>39</sup> I would also include Aristotle and Thomas Aquinas, but since libertarians also claim them, their inclusion would be more controversial.

<sup>40</sup> I think that a person's bad luck is a good reason to mitigate punishment without supposing that the person lacks moral responsibility. So, with respect to punishment, I would fall between Jonathan Edwards, who would not count bad luck as a mitigating factor in the context of punishment, and Derk Pereboom, who holds that, because of the pervasiveness of luck, no one ever deserves any punishment at all.

<sup>41</sup>I benefited from comments by Derk Pereboom, Peter van Inwagen, Katherine Sonderegger, Hilary Kornblith, Gareth B. Matthews, and Maureen Sie. Also, thanks are due to Ishtiyaque Haji and Stefaan E. Cuypers, who commented on an earlier version of this paper at Erasmus University in Rotterdam.