Free Will as an Open Scientific Problem

Mark Balaguer

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Stone Sans and Stone Serif by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Balaguer, Mark.
Free will as an open scientific problem / Mark Balaguer.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 978-0-262-01354-3 (hardcover : alk. paper)
1. Free will and determinism. I. Title.
BJ1461.B25 2010
123'.5—dc22

2009012426

10 9 8 7 6 5 4 3 2 1

1 Introduction

I will argue in this book for a novel view of free will. Most centrally, I will argue that the metaphysically interesting issue in the problem of free will and determinism boils down to a straightforward (and wide open) empirical question about the causal histories of certain neural events. But it will take a long time to get to this conclusion, and along the way, I will argue for several other controversial theses about free will and determinism.

In the present chapter, I will do three things. First, in section 1.1, I will formulate the problem of free will (I begin by giving a traditional formulation of the problem, and then I provide a new and improved formulation). Second, in section 1.2, I provide a few brief remarks about libertarianism. And finally, in section 1.3, I give a quick synopsis of the rest of the book, listing the most important theses that I will be arguing for.

1.1 Formulating the Problem of Free Will

The best way to bring out the problem of free will and determinism is to begin with an old formulation of the problem. I will do this in section 1.1.1, and then in sections 1.1.3–1.1.4, I will explain how and why we have to alter this old formulation of the problem. In between, in section 1.1.2, I will say a bit about compatibilism.

1.1.1 The Old Formulation of the Problem of Free Will

Prior to the emergence of quantum mechanics (or QM) in the early part of the twentieth century, it was extremely easy to bring out the problem of free will and determinism. The problem—or the *traditional* problem, as we can call it—was generated by the fact that people had (or thought they had) prima facie reasons to believe three theses that form an inconsistent set, namely, the following:

(1) Determinism is true (i.e., every event is causally necessitated by prior events together with causal laws);

(2) Human beings have free will; and

(3) Free will is incompatible with determinism.

The prima facie reason for believing (2) was (and is) based on introspection: We have all had the experience of freely choosing from among a number of possible options. The prima facie reason for believing (3), on the other hand, can be put like this:

> If it's really true that all events are causally determined by prior events, then all of my so-called "decisions" and "actions" are determined by prior events, because they are just special kinds of events; indeed, if determinism is true, then it was already determined before I was born that I was going to make all of the decisions I have made during my life; for example, I couldn't have avoided choosing to sit down and write today, because it was already determined before I was born that I *would* choose to sit down and write today. But this is clearly not compatible with free will; we cannot say that I chose to sit down and write of my own free will if it was already determined before I was born that I was going to do this.

Finally, the prima facie case for (1) probably seemed just as obvious to pre-QM philosophers. One might have mounted an argument here in something like the following way:

> Suppose that we strike two billiard balls, A and B, and that A rolls 12 inches before stopping, whereas B rolls 12.1 inches before stopping. Prima facie, it seems that there must be some reason *why* B went farther. Perhaps it was struck a bit harder; or perhaps there was a bit less friction on the part of the table it rolled over; or whatever. We might not know the cause, but it seems that there must have been some cause, because experience tells us that things don't just happen. Prima facie, it seems that everything that happens in the physical world has a cause and, hence, that determinism is true. And this applies to human decisions as well as to any other kind of event. Suppose that I have to choose between two options, A and B, and that I choose A; and suppose that the next day, I am presented with the same choice, and I choose B. It seems that something must have been different about my mental state in the two cases. Perhaps my desire for A decreased in the interim; or perhaps my desire for B increased; or perhaps I just wanted some variety and so I chose B because I remembered choosing A the day before. Or whatever. We might not be able to figure out how my mental state changed, but prima facie, it seems that something about my mental state *must* have changed, because it seems that if I were in the exact same mental state—that is, if I had the same beliefs, desires, fears,

and so on, all of the same strengths—then I would have made the same choice. This is because our decisions flow out of our mental states, that is, our beliefs and desires and fears and so on. In other words, it seems that our decisions are determined by our mental states. And so it seems that our decisions are every bit as determined as other kinds of events are.

It is worth noting that this prima facie argument for (1) does not assume materialism about the mind–brain. If our decisions are determined by our mental states in the way suggested by the above argument, then the decisions of Cartesian souls are every bit as determined as the decisions of materialistic minds, that is, minds that supervene on brains. Note, however, that for those pre-QM philosophers who endorsed both mind–brain materialism and causal determinism about the physical, there was a very obvious argument for the thesis that our *decisions* are determined. One might have put the argument like this: "All physical objects move around according to strictly deterministic laws of motion, like little billiard balls; but human brains are just made up of little physical objects, and human decisions are just physical events, in particular, brain events; therefore, our decisions are wholly determined by the past together with the laws of physics. Thus, given the initial state of the physical world billions of years ago, and given the laws of motion, it was already determined billions of years ago that I was going to choose to write today."

This, then, was the traditional problem of free will and determinism: We had prima facie reasons to believe three theses that, together, form an inconsistent set. Accordingly, there were three possible solutions to the problem: One could reject either (1), (2), or (3). Traditionally, those who rejected (1) usually endorsed (2) and (3) and maintained that human beings possess an indeterministic sort of free will; such philosophers were called *libertarians*. Likewise, those who rejected (2) usually endorsed (1) and (3) and maintained that human beings do not have free will because their actions and decisions are fully determined; such philosophers were called *hard determinists*. And lastly, those who rejected (3) usually endorsed (1) and (2) and maintained that human beings possess free will despite the fact that they are deterministic creatures; such philosophers where called *soft determinists*. (Another term that has sometimes been used interchangeably with 'soft determinism' is 'compatibilism'. Today, however, it is worth keeping these terms separate; compatibilism is the view that (3) is false, whereas soft determinism is the view that (3) is false *and that (1) is true*. Advocates of both views almost always hold that (2) is true, but as we'll see below, many contemporary compatibilists would not want to commit to the truth of (1)—i.e., to the truth of

determinism—and so it would be unwise to use 'compatibilism' and 'soft determinism' interchangeably.)

In sections 1.1.3 and 1.1.4, I will explain how and why the traditional formulation of the problem of free will needs to be altered, and in doing this, I will have a good deal to say about the libertarian view that begins by rejecting thesis (1). Before I get into this, however, I want to say a few words about the compatibilist view that (3) is false.

1.1.2 Compatibilism and the Rejection of (3)

Compatibilism can be defined as the view that free will is compatible with determinism—that is, as the claim that (3) is false. Compatibilist views go back at least to Hobbes, and they have been endorsed by many people since then.[1] Probably the most famous statement of compatibilism was given by Hume. He argued that free will is compatible with determinism by providing an analysis of the notion of free will that, unlike libertarian analyses, doesn't involve indeterminism. Hume's analysis is very commonsensical; in a nutshell, and somewhat roughly, he takes free will to be the capacity to *do what you want*. Hume put it like this (1748, 104):

> 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 move, we also may.

Putting this into contemporary lingo—and altering it somewhat—we arrive at the following definition:

> A person *S* has *Humean freedom* iff *S* is capable of acting in accordance with his or her choices, and of choosing in accordance with his or her desires; that is, iff it is the case that *if* he or she chooses to do something then he or she does it, and if (all things considered) he or she wants to make some choice then he or she does make that choice.

Hume's argument for compatibilism is that (a) Humean freedom captures the ordinary notion of free will (that is, Humean freedom *is* free will), and therefore (b) free will is compatible with determinism. The only controversial claim here is (a). The inference from (a) to (b) is entirely trivial, because Humean freedom is clearly compatible with determinism. Humean freedom requires only that our actions flow out of our decisions and our decisions flow out of our desires; but this could be the case even if all of our desires and decisions and actions are causally determined. Suppose that (i) Ralph is in some mental state *M*, including some particular ensemble of desires; and (ii) Ralph's being in *M* was causally determined by prior

events (indeed, by events that occurred before Ralph's birth); and (iii) Ralph's being in *M* deterministically causes him to make some decision *D*, and this in turn deterministically causes him to perform some action *A*. Then *D* counts as a Humean free decision and *A* counts as a Humean free action (because Ralph is acting here in accordance with his decision and deciding in accordance with his desires); but, of course, *D* and *A* are also causally determined by prior events. Therefore, the claim that human beings are Humean free is perfectly compatible with the thesis that determinism is true, and so again, the only controversial part of Hume's argument is the thesis that free will is in fact Humean freedom, that is, the thesis that the notion of Humean freedom captures the ordinary notion of free will.

Humean freedom is not the only compatibilist notion of freedom in the literature; various other philosophers—P. F. Strawson (1962), Frankfurt (1971), Watson (1975), Dworkin (1988), Wolf (1990), Double (1991), Fischer (1994), Fischer and Ravizza (1998), Wallace (1994), Mele (1995), and Bok (1998), to name just a few[2]—have offered alternative compatibilist definitions of free will (and/or moral responsibility). It is important to note, however, that like Humean freedom, all of these other compatibilist notions of freedom (or at any rate, all the ones that I'm aware of) are not just compatible with determinism, but *obviously* so. For instance, to say just a few words about one of these cases, *Frankfurtian freedom* can be defined (very briefly) as the ability to control, with second-order attitudes, which of your first-order desires will affect your behavior; but given this, it seems that Frankfurtian freedom is clearly compatible with determinism, because one could be causally determined to have (and be governed by) a second-order desire *d* for some first-order desire *e* to control one's behavior (i.e., to take precedence over any conflicting first-order desire *f*). The same point can be made in connection with all of the other compatibilist notions of freedom in the literature, but I won't argue this here, because in each case the point is just as obvious and uncontroversial as it is in connection with Humean freedom and Frankfurtian freedom. In other words, in connection with each of these different kinds of freedom, the controversial question is not whether it is indeed compatible with determinism, but whether it provides a correct analysis of the notion of free will—that is, whether it *is* free will.

It perhaps goes without saying that a similar point can be made about compatibilism in general: Compatibilism is a controversial view because many people think that the correct analysis of free will is given by the notion of libertarian freedom, which, as we'll see, requires indeterminism in a completely transparent way. I will have much more to say about libertarian freedom below, in sections 1.1.3, 1.1.4, and 1.2.

It is also worth noting here that just as it's obvious that the various compatibilist kinds of freedom are in fact compatible with determinism, it is also obvious (in at least most of these cases) that human beings do in fact possess these kinds of freedom. For instance, it's obvious that we are Humean free, because we are clearly capable of acting in accordance with our decisions and deciding in accordance with our desires. Likewise, it's equally obvious that we possess Frankfurtian freedom, because we clearly have the ability to want certain first-order desires to affect our behavior, and at least sometimes, we have the ability to act on these second-order wants. The same point can be made in connection with just about all of the compatibilist varieties of freedom in the literature. I won't say that this point is *completely* obvious in connection with *all* of these kinds of freedom, because one might think there are a few problematic issues here. For instance, some compatibilist kinds of freedom require something like reasons responsiveness (see, e.g., Fischer and Ravizza 1998), and as we'll see in chapter 4, one might think that various psychological studies suggest that our actions are less responsive to our reasons than we might have thought. But in the end, it still seems pretty clear that at least sometimes, and probably very often, our actions are responsive to our reasons in obvious and important ways.

Finally, it's worth noting that in the past (before the emergence of quantum mechanics), most compatibilists endorsed determinism. Again, such philosophers were known as soft determinists. Some of these philosophers (Hobart for sure and arguably Hobbes and Hume as well) also held that freedom *requires* determinism. But today, very few compatibilists would want to commit to either of these claims. Most compatibilists would likely say that they have no idea whether determinism is true (this, they might say, is a question for physicists) but that as far as the issue of free will is concerned, it just doesn't matter whether it's true, because free will doesn't require determinism or indeterminism.

1.1.3 An Intermediate Formulation of the Problem of Free Will

The emergence of QM undermined the traditional way of understanding the problem of free will by undermining our prima facie reasons for believing determinism and, indeed, by revealing that determinism is not the sort of doctrine that can be motivated by prima facie, pretheoretic, armchair arguments. Rather, it is a controversial empirical thesis about the workings of the physical world; in short, it is a question for scientists, most notably physicists. Today, many (perhaps even most) physicists and philosophers of physics reject determinism and endorse indeterminism. And just about everyone would agree that the question is controversial—

that at present, we do not have decisive evidence for either determinism or indeterminism. (In chapter 4, I will argue for a much stronger claim; I will argue that, in fact, we have no good reason whatsoever for endorsing determinism or indeterminism.)

So, given that it no longer seems correct to claim that we have prima facie reasons to believe determinism, we can no longer formulate the philosophical problem of free will in the traditional way. But it turns out that the problem of free will can be separated from its traditional formulation; essentially the same problem (or at least a very similar problem) can be motivated without claiming that we have reason to believe determinism. In short, what we have to do here is replace the appeal to determinism, in the traditional formulation of the problem, with another thesis, one that (a) still gives rise to a traditional sort of worry about free will (just as determinism does), but (b) unlike determinism, has some prima facie support, or plausibility.

How can we come up with such a thesis? Well, we can start by noticing that the traditional worry about free will is really a worry about the existence of an *indeterministic* sort of free will—or, more precisely, a *libertarian* sort of free will—and by reflecting on what libertarianism is supposed to involve. In particular, the point I want to bring out here is that any interesting variety of libertarian free will is going to involve more than just indeterminism. At the very least, it is also going to involve some sort of *appropriate nonrandomness*.

The notion of appropriate nonrandomness is going to be crucially important in what follows. Different people might give different accounts of what this sort of nonrandomness consists in, but the basic idea is that in order for a decision to count as appropriately nonrandom (and hence free), the agent in question has to be centrally involved in the decision. Probably the most standard thing to say here—and I will go along with this—is that appropriate nonrandomness consists most importantly (but perhaps not entirely) in the given agent *authoring and controlling* the decision; that is, it has to be *her* decision, and she has to control which option is chosen. I will discuss this in more detail in chapter 3; I will also discuss some other possible requirements for appropriate nonrandomness, most notably, rationality and what Kane (1996) has called *plural* authorship, control, and rationality; but for now, these brief remarks are good enough.

It's easy to see why any interesting variety of libertarian free will is going to require not just indeterminism, but some sort of appropriate nonrandomness as well. To appreciate this point, imagine that Martians have implanted a chip in someone's head (say, Sylvia) and are controlling her choices and actions via remote control; but imagine that, sometimes, environmental

noise garbles the signal from the remote control and that when the signal reaches Sylvia's head, it causes her to choose not as the Martians wanted her to choose, but in a different way; and finally, imagine that a specific signal arrives at Sylvia's head, that it's partially garbled, and that it causes Sylvia to choose as the Martians wanted her to choose but that this result was not causally determined, i.e., that the partially garbled signal might just as well have caused her to choose differently. In this scenario, it seems pretty clear that Sylvia's decision was not free because it wasn't *her* making the decision—or if you'd rather, because she didn't author or control the decision. One might put the point here by claiming that in this scenario, it seems more accurate to say that the decision *happened* to Sylvia, or was inflicted on her, than that she chose of her own free will.

It seems to follow from these considerations that some sort of appropriate nonrandomness is required for genuine free will. The ordinary notion of free will may or may not require indeterminism—that is a controversial question. But regardless of what we say about this, it seems clear that the ordinary notion does require some sort of agent-involving nonrandomness; this is why we have the intuition that in the above scenario, Sylvia's decision is not free. And this should hardly be surprising. After all, the sort of nonrandomness that we're talking about here is really just a sort of *agent-involvedness*, and it should be pretty clear that we cannot say that a person chose of her own free will if she was not relevantly involved in the given decision—that is, if the decision wasn't *hers*, or she didn't control which option was chosen, or some such thing. This just seems obvious.

Finally, given that some sort of appropriate nonrandomness is required for ordinary free will, it seems to follow that any interesting variety of libertarian freedom will require appropriate nonrandomness as well. Indeed, I think we can say that if libertarian freedom didn't require some sort of agent-involving nonrandomness, then it would be a pretty silly, uninteresting variety of freedom. But if libertarian freedom requires indeterminacy *and* appropriate nonrandomness, then it's easy to formulate a hypothesis that (a) is weaker than determinism but (b) still generates a worry about the existence of libertarian free will. The thesis I have in mind is simply this:

> *Determined-or-Randomism (D-or-R-ism)*: None of our decisions is both undetermined and appropriately nonrandom; that is, all of our decisions are either (i) causally determined by prior events or (ii) random in the sense that they're not appropriately nonrandom.

One idea, then, would be to replace the appeal to determinism, in the traditional formulation of the problem of free will, with an appeal to D-or-R-

ism. If we did this, then the resulting problem of free will would be that we have prima facie reasons to believe the following three (jointly inconsistent) theses:

(1') D-or-R-ism is true;

(2) Human beings have free will; and

(3') Free will is incompatible with D-or-R-ism.

The prima facie reason to believe (2) is the same as it was before. Moreover, the prima facie reason to believe (3') has really already been given as well. For I just argued that free will requires appropriate nonrandomness, and if we combine this with the above prima facie reason to believe (3)—that is, the prima facie reason to believe that free will is incompatible with determinism—then we arrive at the result that free will is incompatible with both determinism and randomism and, hence, that it's incompatible with D-or-R-ism. (It is worth noting here that since it is more or less uncontroversial that free will requires some sort of appropriate nonrandomness, the only really controversial part of (3') is the claim that free will is incompatible with determinism. Thus, one might think that in formulating this new version of the problem of free will, we should just stick with a modified version of (3), claiming that we have prima facie reason to believe the following:

(3) Free will is incompatible with determinism [and hence also with D-or-R-ism].)

Finally, there are also prima facie reasons—or at least one such reason—to believe (1'). The main argument here has been advanced by a few different philosophers, including Hobbes (1651), Hume (1748), and Hobart (1934). We can put the argument like this:

> Indeterminism seems to entail a kind of randomness. It seems that if an event is undetermined, then it's uncaused and, hence, accidental. In other words, it *just happens*, or it happens randomly. Therefore, prima facie, it seems that if our decisions are undetermined, then they are random, and so they couldn't possibly be "appropriately nonrandom." Or to put the point the other way around, if our decisions are appropriately nonrandom, then *we* author and control them, or something along these lines, and this, it seems, can only mean that *we determine what we choose and don't choose*, presumably for rational reasons; thus, it seems that if our decisions are appropriately nonrandom, then they couldn't possibly be undetermined. Therefore, prima facie, it seems impossible for a decision to

be undetermined and appropriately nonrandom at the same time, and so it seems that D-or-R-ism is true; that is, it seems that all of our decisions are either determined or random in the sense of being not appropriately nonrandom.

(Notice that while this argument involves the claim that indeterminacy *entails* a lack of appropriate nonrandomness, D-or-R-ism itself does *not* involve this claim. D-or-R-ism just says that, in fact, all of our decisions *are* either determined or random.)

In any event, given the above remarks, it seems that there are at least prima facie reasons to believe (1'), (2), and (3'). Thus, this gives us an alternative formulation of the problem of free will.

1.1.4 The Final (or "New and Improved") Formulation of the Problem of Free Will

But this isn't yet the best way to formulate the problem of free will. The trouble is that D-or-R-ism is still too strong. It is of course weaker than determinism, and so we've made some progress here, but we can do better. In other words, it's possible to generate a traditional sort of worry about indeterministic, libertarian free will by appealing to a hypothesis that's even weaker than D-or-R-ism. And as we'll presently see, when we do this, we arrive at a much better formulation of the problem.

We can locate the hypothesis I have in mind by noticing that we haven't yet captured everything that's required for libertarian free will. In order for a decision to be genuinely free in the libertarian sense, it's not enough for it to be undetermined and appropriately nonrandom. It also needs to be the case that the indeterminacy is *relevant* to the nonrandomness. Thus, I want to define libertarianism as follows:

> *Libertarianism* is the view that human beings are L-free, where a person is *L-free* if and only if she makes at least some decisions that are such that (a) they are both undetermined and appropriately nonrandom, and (b) the indeterminacy is relevant to the appropriate nonrandomness in the sense that it *generates* the nonrandomness, or *procures* it, or *enhances* it, or *increases* it, or something along these lines.[3]

(One might want to add another thesis to libertarianism, namely, the thesis that the notion of L-freedom provides an adequate definition of the ordinary notion of free will. If we did this, then libertarianism would entail (i) that human beings have free will and (ii) that free will is incompatible with determinism. This is probably the "standard" way to define

libertarianism, but I don't want to set things up this way. I will explain why in section 1.2, but for now, I just want to ignore this issue.)

It's easy to see why clause (b) is needed in the definition of libertarianism. Without this, some decidedly nonlibertarian views would count as versions of libertarianism. Consider, for instance, the following view:

> ($) Hume was essentially right about what free will is; in particular, free will consists essentially in the ability to act and choose in accordance with your desires. Moreover, we humans do in fact have free will: we are capable of choosing in accordance with our desires because most of our choices are in fact *caused by* our desires. But some aspects of Hume's view are *not* right. In particular, it turns out that our decisions are not determined, because there are various kinds of quantum indeterminacies in our decision-making processes. But our decisions are still, so to speak, "for-all-practical-purposes determined." More specifically, they are probabilistically caused—by our desires—with a high degree of certainty. For instance, suppose that some agent (say, Smith) has to choose between two options, A and B; and suppose in addition that Smith's desires, or reasons for choosing, clearly favor option A; and finally, suppose that Smith does in fact choose A. Then according to the present view, just prior to the moment of choice, the probability of Smith choosing A was very, very high (say, .9999999), but it was not 1. So our decisions are not determined, but this doesn't really matter, because they're still for-all-practical-purposes determined by our desires. Thus, our decisions are still appropriately nonrandom and free.

If we didn't include clause (b) in the definition of libertarianism, then ($) would count as a version of libertarianism, because it entails that our decisions are both undetermined and appropriately nonrandom. But ($) clearly *isn't* a version of libertarianism; it's an essentially Humean view with an unimportant smidge of indeterminism thrown in. One way to appreciate this point is to notice that according to ($), we have free will *despite* the fact that indeterminism is true. Libertarians, on the other hand, think that we have free will *because* indeterminism is true; they think that indeterminacy is *needed* for free will—or at any rate, they think it's needed for the kind of free will that they have in mind. And this is why it has to be built into the definition of libertarianism that the indeterminacy in question *generates* the nonrandomness (or enhances it, or increases it, or some such thing).

Given that libertarianism should be defined in the above way, it follows that we can weaken the thesis that we need to appeal to in order to generate

a problem for free will. We don't need to appeal to determinism or even to D-or-R-ism. All we need to appeal to here is the following thesis:

> *No Freedom-Enhancing Indeterminism (or for short, FE-determinism)*: There are no freedom-enhancing indeterminacies (i.e., no freedom-enhancing undetermined events) in any human decision-making processes. In other words—and more precisely—human beings do not possess L-freedom; that is, libertarianism is false.

It is worth noting here that FE-determinism is similar in a certain way to determinism. In particular, it's just a narrowed-down version of determinism. Determinism says that there are no undetermined events of any kind anywhere in the universe. FE-determinism, on the other hand, says that there are no undetermined events *in our decision-making processes that generate or increase appropriate nonrandomness*. So determinism and FE-determinism are similar sorts of theses. The difference is that FE-determinism is a much narrower claim that zeros in on the kinds of indeterminacies that might be relevant to the issue of free will. And the benefit of narrowing the claim down in this way, as we'll presently see, is that when we do this, we arrive at a thesis that we seem to have prima facie reasons to believe.

One might put forward a few different prima facie arguments for FE-determinism. The one I'm going to describe here is similar to the Hobbes-Hume-Hobart argument for D-or-R-ism, except that it's weaker and hence more plausible. The Hobbes-Hume-Hobart argument proceeds by claiming that indeterminacy entails a lack of appropriate nonrandomness. The argument for FE-determinism is similar, but it doesn't commit to the thesis that indeterminacy automatically destroys appropriate nonrandomness; it leaves open the possibility that there might be some indeterminacies in our decision-making processes that are completely irrelevant to the nonrandomness of our decisions and, hence, that don't diminish this nonrandomness. What the argument claims, though, is that there couldn't be any indeterminacies that *increase* appropriate nonrandomness. One might formulate the argument here in something like the following way:

> It's hard to see how the introduction of an undetermined event into a decision-making process could *increase* the authorship or control that the agent in question has over the decision. It seems (prima facie) that at best, an undetermined event could be irrelevant to the agent's authorship and control. Authorship, control, and nonrandomness have to do with it being the case that it was the *agent* who made the decision, or who determined which option was chosen. But given this, it's hard to see how the insertion of an undetermined event could help. It seems that if we insert an unde-

> termined event into a decision-making process, this would be to insert an element of randomness; so it seems that this would either (a) undermine the appropriate nonrandomness of the decision or, at best, (b) leave the level of appropriate nonrandomness alone. How could the insertion of a *random* element increase appropriate *non*randomness? How could this make it the case that it was the *agent* who performed the given action, or made the given decision? It's hard to see how it could.

(It's important to note that this is just one way to formulate a prima facie argument for FE-determinism—or, equivalently, against libertarianism. When we get to chapter 3, we will encounter some more detailed objections to libertarian views, e.g., the luck objection and the objection that libertarianism leads to an unacceptable regress. For now, though, the above prima facie argument is good enough.)

Given all of this, we can replace the appeal to determinism, in the traditional formulation of the problem of free will, with an appeal to FE-determinism. If we do this, we arrive at a new and improved version of the problem of free will, one that's generated by pointing out that we have—or seem to have—prima facie reasons to believe the following three (mutually inconsistent) theses:

(1") FE-determinism is true (i.e., libertarianism is false).

(2) Human beings have free will.

(3) Free will is incompatible with determinism (and, hence, with FE-determinism).

Now, it might seem that thesis (3) should be replaced, in the new and improved version of the problem, with the following:

(3") Free will is incompatible with FE-determinism.

But I want to argue that free will is incompatible with FE-determinism if and only if it's incompatible with determinism. (If this is right, then the question of whether (3") is true reduces to the question of whether free will is compatible with determinism; and given this, it makes sense to favor (3) over (3") for the simple reason that it enables us to preserve some traditional lingo, couching the important issue in terms of the question of whether compatibilism is true, i.e., the question of whether free will is compatible with *determinism*.) What, then, is the argument for the claim that free will is incompatible with FE-determinism if and only if it's incompatible with determinism? Since this is a biconditional, we can justify it by arguing for two conditional claims, namely, (a) if free will is

incompatible with determinism, then it's incompatible with FE-determinism, and (b) if free will is incompatible with FE-determinism, then it's incompatible with determinism—or equivalently, if free will is compatible with determinism, then it's compatible with FE-determinism. Both of these conditionals are easy to justify.

Let's start with (b). Suppose free will is compatible with determinism. Then it's obviously compatible with FE-determinism—that is, with the claim that there are no freedom-enhancing indeterminacies in any of our decision-making processes—for the simple reason that it's compatible with the claim that there are no indeterminacies (i.e., no undetermined events[4]) of any kind anywhere in the universe.

Let's move on now to claim (a). Suppose that free will is incompatible with determinism. Then free will requires indeterminism. But it seems clear that if indeterminacies are required for free will, it's because they're required for appropriate nonrandomness. (I can't imagine why else indeterminism would be needed here; the whole worry about free will and determinism is that if our decisions are causally determined by prior events, then they aren't made by *us*, i.e., they aren't authored or controlled by us and, hence, aren't appropriately nonrandom.) But if indeterminism is required for appropriate nonrandomness, then it seems to follow that free will requires a variety of indeterminism that procures or increases appropriate nonrandomness—that is, it requires a freedom-enhancing indeterminism of precisely the kind that's ruled out by FE-determinism. But, of course, this is just to say that free will is incompatible with FE-determinism, that is, with the claim that there are no freedom-enhancing indeterminacies in any of our decision-making processes. And so it follows that claim (a) is true: if free will is incompatible with determinism, then it's also incompatible with FE-determinism.

(A second, perhaps ultimately equivalent, way to motivate thesis (a) is to point out that the libertarian notion of free will is really the only reasonable incompatibilist notion of free will out there [at any rate, I can't see how incompatibilists have any other reasonable options]. Given this, it seems very plausible to suppose that if free will is incompatible with determinism, it's because free will is what libertarians say it is—which is just to say that free will is incompatible with the falsity of libertarianism, or equivalently, with the truth of FE-determinism.)

Given (a) and (b), it follows that (3") is essentially equivalent to (3). Thus, for the sake of preserving some traditional lingo, I want to formulate the problem of free will in terms of thesis (3) and the issue of the incompatibility of free will and *determinism*.

So, again, the (new and improved) problem of free will and determinism is that we have prima facie reasons to believe the following three (mutually inconsistent) theses:

(1") FE-determinism is true (i.e., libertarianism is false).

(2) Human beings have free will.

(3) Free will is incompatible with determinism (and, hence, with FE-determinism).

As with the traditional problem, there are three possible solutions to this new and improved version of the problem. In particular, we can reject either (1"), (2), or (3). Again, those who reject (3) are called *compatibilists*; usually these philosophers will also endorse (1") and (2); we can call these people *soft FE-determinists*. Likewise, those who respond to the problem by rejecting (2) will usually endorse (1") and (3)—this, of course, is why they reject (2); we can call these people *hard FE-determinists*.[5] And, of course, those who reject (1") are called *libertarians*. Before going on, I want to say a bit more about this view.

1.2 Some Remarks on Libertarianism

I defined libertarianism above as the negation of (1"). To repeat, in my lingo,

> *Libertarianism* is the view that human beings are L-free, where a person is *L-free* if and only if she makes at least some decisions that are such that (a) they are both undetermined and appropriately nonrandom in the sense discussed above, and (b) the indeterminacy is *relevant* to the appropriate nonrandomness in the sense that it generates it, or procures it, or increases it, or some such thing.

I should note here that other philosophers might define libertarian freedom slightly differently. For instance, Kane (1996, chapters 3 and 5) defines it in terms of an agent having ultimate responsibility for, or "sole authorship" of, his or her decisions. But as far as I know, the various definitions of libertarian freedom (Kane's included) all require indeterminism, and they all require the indeterminism to be an important factor in procuring some sort of appropriate nonrandomness, where this involves (at the very least) something like authorship and control.

Nonetheless, despite this similarity that runs through all versions of libertarianism, it should be noted that there are many varieties of libertarian-

ism, because the thesis that we humans are L-free can be developed in a number of different ways. For instance, whereas many early libertarians endorsed mind–brain dualism, many contemporary libertarians are mind–brain materialists. And whereas some libertarians endorse the idea that there is such a thing as irreducible agent causation, others reject this idea and maintain that all causation reduces to event causation. Moreover, among those libertarians who reject irreducible agent causation, there is a distinction to be drawn between event-causal libertarians (who hold that our L-free decisions are probabilistically caused by agent-involving events) and noncausal libertarians (who hold that our L-free decisions are uncaused). And finally, some libertarian views are Valerian (i.e., they place the important indeterminacy prior to the moment of choice) and others are non-Valerian (i.e., they place the important indeterminacy *at* the moment of choice). I will say more about these different versions of libertarianism in chapter 3.

The point I want to bring out now, though, is that my definition of libertarianism is a bit nonstandard in a certain way. I have defined libertarianism as the view that human beings are L-free, but I think it is fair to say that many people have used 'libertarianism' to denote the view that human beings are L-free *and that the notion of L-freedom provides a correct analysis of the ordinary notion of free will*. I do not want to define libertarianism in this way for two reasons. First, this definition seems unpleasing to me, because it leaves out a view that I think is best thought of as a version of libertarianism. Suppose someone held that (a) human beings do possess L-freedom (and that this is an important fact about the nature of human decision making) but (b) ordinary-language utterances of 'free will' refer not to L-freedom but to some compatibilist sort of freedom (say, Humean freedom). It seems to me that this view is best thought of as a variety of libertarianism, and so I do not want to include in the definition of libertarianism the claim that L-freedom captures the ordinary notion of free will. (By the way, I would not say that the traditional definition of libertarianism is *wrong* in this respect, because 'libertarianism' is a theoretical term of art and so it is defined by stipulation; definitions of ordinary-language terms can be wrong because they can fly in the face of ordinary usage and intentions [and perhaps for other reasons as well—see chapter 2 for more on this]; but this doesn't seem to be the case with definitions of words like 'libertarianism'.)

My second reason for favoring my definition of libertarianism over the more traditional one is that it clarifies things by keeping the metaphysical thesis that human beings are L-free separate from the (entirely independent) semantic thesis that the ordinary notion of free will is accurately

defined by the notion of L-freedom. Now, I suppose we could achieve even more clarity here by introducing two different terms; for example, we could define *metaphysical libertarianism* as the view that human beings are L-free and *semantic libertarianism* as the view that the notion of L-freedom provides a correct definition of 'free will'. But there is no need for me to use these terms here, because I am not going to be very concerned with semantic libertarianism; for the most part, I will be concerned only with metaphysical libertarianism—or as I will be calling it, *libertarianism*.

So those are my reasons for leaving the above semantic thesis out of the definition of libertarianism. But, of course, nothing important is going to turn on this terminological point.

Now, given that libertarianism (as I'm defining it here) does not involve the claim that free will is L-freedom, it follows that libertarianism is consistent with compatibilism. There is no inconsistency in claiming that (a) human beings possess L-freedom, but (b) L-freedom does not provide a good analysis of the ordinary notion of free will and, in fact, that notion is best analyzed in some compatibilistic way. Now, this might seem like an odd result—that libertarianism and compatibilism are consistent—but of course it's just a trivial result of the fact that I've defined libertarianism in a somewhat nonstandard way. Nonetheless, while there is nothing particularly interesting about the fact that claims (a) and (b) are consistent, there is another point lurking in the background here that I think is more interesting and important. The point I have in mind is that very few people have simultaneously endorsed these two views. This strikes me as a bit odd. Indeed, it seems more than odd; it seems *fishy*. Many people think that human beings are L-free, and many think that the ordinary notion of free will is compatible with determinism; these two theses are perfectly consistent with one another, and indeed, they are entirely independent of each other—one of them is about human decision-making processes and the other is about the meaning of a certain ordinary-language expression. So why don't more people endorse these two theses together? The answer, I think, is that philosophers have let their views on these two issues influence each other in irrational ways (for instance, I think that many people have endorsed compatibilism more or less *because* they think human beings aren't L-free, and I think that many people have endorsed libertarianism *because* they endorse incompatibilism). I will not return to this point in the present book, but in chapter 2, I will provide more reason for thinking that the two issues here are entirely independent of one another, and it will follow from this that there is simply no good reason to think that we shouldn't simultaneously endorse libertarianism (as I'm defining it here) and compatibilism.

(Indeed, as I pointed out above, this is one of my reasons for favoring my definition of libertarianism—because it helps us to keep separate two theses that are in fact completely independent of one another.)

In any event, even if libertarianism (as I've defined it here) is compatible with compatibilism, it is worth noting that almost all compatibilists reject libertarianism (as I define it), and vice versa. And again, we can call those compatibilists who reject libertarianism *soft FE-determinists*. And if we like, we can call libertarians who reject compatibilism *metaphysical and semantic libertarians*—or perhaps *incompatibilistic libertarians*. Finally, it's worth noting that virtually all the proponents of both of these views endorse thesis (2)—that is, the thesis that human beings have free will. Most libertarians endorse (2) because they think that humans are L-free and that this is free will. And most compatibilists endorse (2), because, as we saw above, human beings pretty obviously possess just about all of the standard varieties of compatibilist freedom.

1.3 Synopsis of the Book

As I've set things up, the (new and improved) problem of free will is generated by the fact that we have prima facie reasons to believe the following three theses:

(1") FE-determinism is true;

(2) Human beings have free will; and

(3) Free will is incompatible with determinism (and hence with FE-determinism).

But prima facie reasons are just that—prima facie reasons. The challenge is to figure out which of these three theses are really true. Thus, if we want to *solve* the problem of free will, we have to ask the following three questions:

The FE-determinism question: Is FE-determinism true?

The do-we-have-free-will question: Do human beings have free will?

The compatibilism question: Is free will compatible with determinism (and hence with FE-determinism)?

But I will argue in chapter 2 that it's possible to get more focused with these questions. In particular, I will argue that these three questions reduce to two other questions, namely,

The what-is-free-will question: What is free will? (We can take this as being equivalent to the question 'What is the correct analysis of the notion of

free will?' and also to the question 'What is the correct definition of the term "free will"?' But we cannot assume without argument that these questions are solely about folk meaning, or ordinary-language usage and intentions; I will discuss this issue in chapter 2.)

and

The which-kinds-of-freedom-do-we-have question: Which kinds of freedom do humans have? That is, do they have L-freedom?; and do they have Humean freedom?; and do they have Frankfurtian freedom?; and so on. (Actually, to be more precise, we can formulate this question as asking which kinds of "freedom" humans have, since some or all of the kinds of "freedom" we're asking about here might fail to *be* free will, according to the correct answer to the what-is-free-will question.)

I will argue that if we could answer these latter two questions, then we could thereby answer the first three questions. More specifically, I will argue that (i) the compatibilism question reduces to the what-is-free-will question; and (ii) the FE-determinism question reduces to (or collapses into) the which-kinds-of-freedom-do-we-have question; and (iii) the do-we-have-free-will question collapses into the what-is-free-will question *and* the which-kinds-of-freedom-do-we-have question. In addition, I will also argue that (iv) while the what-is-free-will question is definitely relevant to the do-we-have-free-will question in a certain sort of way, it is not relevant to that question in any nontrivial or metaphysically interesting way; indeed, I will argue that the what-is-free-will question is irrelevant to *all* substantive questions about the nature of human decision-making processes, except in a trivial way. Thus, from points (i) and (iv), it follows that (v) the compatibilism question is likewise irrelevant to substantive questions about the nature of human decision-making processes (most notably, the do-we-have-free-will question), except in a trivial way. And from points (iii) and (iv), it follows that (vi) the only metaphysically interesting question that has any bearing at all on the do-we-have-free-will question is the which-kinds-of-freedom-do-we-have question. Finally, I will also argue that the which-kinds-of-freedom-do-we-have question boils down largely (though perhaps not entirely) to the libertarian question (i.e., the question of whether human beings are L-free); and when we combine this with thesis (vi), it gives us the result that (vii) the metaphysical issue inherent in the problem of free will reduces largely (though, again, perhaps not entirely) to the libertarian question.

Given the way I've set things up here, it seems to make sense to think of the problem of free will as being *constituted by* the three questions listed

above—that is, the FE-determinism question, the do-we-have-free-will question, and the compatibilism question. And given the reductive conclusions mentioned in the preceding paragraph, we might go on to say that the problem of free will is, in the end, *really* constituted by the what-is-free-will question and the which-kinds-of-freedom-do-we-have question. But in fact, there's another question that's relevant to philosophical discussions of free will that *doesn't* reduce to the what-is-free-will question and/or the which-kinds-of-freedom-do-we-have question, namely,

> *The moral responsibility question*: Which kinds of freedom (or "freedom") are required for moral responsibility?

I think that if we could answer this last question in addition to the what-is-free-will question and the which-kinds-of-freedom-do-we-have question, then we could answer just about all of the important philosophical questions in the area of free will. But I won't argue this point here. Instead, I will argue (still in chapter 2) that (viii) like the compatibilism question and the what-is-free-will question, the moral responsibility question is essentially irrelevant to substantive questions about the metaphysics of human free will.

Because I argue in chapter 2 for thesis (vi) above—that is, the thesis that the which-kinds-of-freedom-do-we-have question is the only metaphysically interesting question that's relevant to the do-we-have-free-will question—and because my central concern in this book is the metaphysics of human free will and not the semantics of the term 'free will', most of the rest of the book is concerned with the which-kinds-of-freedom-do-we-have question, and in particular, with the libertarian question, which, again, is the most important and controversial subquestion of the which-kinds-of-freedom-do-we-have question. But before moving on to the libertarian question, I say a few words, at the end of chapter 2, about how we might answer the what-is-free-will question, the compatibilism question, and the moral responsibility question.

In chapter 3, I turn to the libertarian question. It is widely believed that libertarianism (i.e., the view that human beings are L-free) could not be true. One argument for this, sketched above, is based on the idea that even if there are various kinds of causal indeterminacies in our decision-making processes, these indeterminacies could not increase or procure appropriate nonrandomness. A few libertarians—see, for example, van Inwagen 1983, Kane 1985, 1996, and Ginet 1990—have responded to worries like this by trying to explain how libertarianism *could* be true, that is, by trying to show that libertarianism is at least possible. In chapter 3, I will argue for a

much stronger conclusion. I will argue that there's a certain category of our decisions (I will call them *torn decisions*, and I'll characterize them in chapter 3) for which the following is true: If they're undetermined in the appropriate way (and I'll be very clear about the sort of indeterminacy that's required here), then they're L-free—that is, (a) they're not just undetermined but also appropriately nonrandom, and (b) the indeterminacy in question increases or procures the appropriate nonrandomness. Notice that my claim here is not simply that if our torn decisions are appropriately undetermined, then they *could* be L-free; I'm going to argue that if they're appropriately undetermined, then they *are* L-free. If I'm right about this (and about a couple of other minor points that I will argue in chapter 3), then the question of whether libertarianism is true—that is, of whether human beings are L-free—just reduces to the question of whether some of our torn decisions are undetermined in the appropriate way.

(It might seem that my thesis here—that if our torn decisions are undetermined in the right way, then they're also appropriately nonrandom and L-free—is inconsistent with something I said in section 1.1.3, when I was discussing the case of Sylvia. Sylvia's decision was undetermined but *not* appropriately nonrandom [and hence not L-free]. But in fact, there is no inconsistency here, because (a) Sylvia's decision was not a *torn* decision, and (b) it wasn't undetermined *in the right way* [and again, I'll characterize torn decisions and the required sort of indeterminacy in chapter 3].)

In developing my arguments in chapter 3, I will also be constructing a novel version of libertarianism. The view will be non-Valerian (i.e., it will place the important indeterminacy at the moment of choice), and it will be entirely naturalistic and event-causal; that is, it will not involve any sort of mind–brain dualism or irreducible agent causation. Thus, combining this with my conclusion that the libertarian question reduces to the question of whether some of our (torn) decisions are (appropriately) undetermined at the moment of choice, we obtain the result that the libertarian question reduces to a straightforward empirical question about the physical world. More specifically, it reduces to a question about the causal histories of the neural events that are our (torn) decisions. In particular, it's just the question of whether any of these neural events are causally undetermined in the appropriate way.

Finally, in chapter 4, I will argue that there are no good arguments on either side of the question of whether some of our torn decisions are undetermined in the appropriate way, that is, the way that's required for L-freedom. Since this is an empirical question, there is, I argue, no real hope that any a priori argument could succeed here. Nonetheless, I begin by saying

what's wrong with a few a priori arguments (or allegedly a priori arguments) that one might attempt here; most notably, I say a few words about what's wrong with the Kant-inspired argument that human beings must be L-free because they're morally responsible for their actions. Then I turn to empirical arguments. I examine the existing evidence in physics and neuroscience and argue that we have no good empirical reason to endorse or reject the thesis that some of our torn decisions are undetermined in the relevant way. I spend most of my time arguing that we have no good empirical reason to *reject* this thesis, because I think it's more or less obvious and uncontroversial that we have no good empirical reason to *endorse* it. Finally, if I'm right that there are no good arguments for or against the relevant sort of indeterminism, then it follows that there are no good arguments for or against libertarianism.

The arguments and conclusions of this book can be seen as fitting into a certain kind of antimetaphysical view. If my arguments here are cogent, then the main philosophical questions about free will reduce to the following three questions:

> *The which-kinds-of-freedom-do-we-have question*: Which kinds of freedom (or "freedom") do humans have? That is, do they have L-freedom?; and do they have Humean freedom?; and do they have Frankfurtian freedom?; and so on.
>
> *The what-is-free-will question*: What is free will?
>
> *The moral responsibility question*: Which kinds of freedom (or "freedom") are required for moral responsibility?

In addition, if my arguments are correct, we get the following results: (a) the metaphysically interesting issue inherent in the problem of free will reduces to the which-kinds-of-freedom-do-we-have question; and (b) the which-kinds-of-freedom-do-we-have question reduces largely to the libertarian question; and (c) the libertarian question reduces to a wide open empirical question about whether our torn decisions are (appropriately) undetermined at the moment of choice. Moreover, in connection with the use of the word 'largely' in point (b), I think it can also be argued—though I won't really argue it here—that if there are any other controversial subquestions of the which-kinds-of-freedom-do-we-have question, aside from the libertarian question, then they too boil down to empirical questions about human decision-making processes. And if this is right, then it follows that the metaphysically interesting issue inherent in the problem of free will reduces to straightforwardly empirical questions about us and our decision-making processes.

What about the what-is-free-will question and the moral responsibility question? Well, if my arguments are correct, then we know that these two questions aren't relevant (in any nontrivial way) to metaphysical questions about the nature of human beings, for example, the do-we-have-free-will question. But what *are* they relevant to? What are these two questions *about*? Well, as we'll see in chapter 2, there are multiple views one might endorse here. One view is that the what-is-free-will question and the moral responsibility question are straightforward empirical questions about the ordinary-language meanings of 'free will' and 'morally responsible'—that is, about the usage and intentions associated with these two expressions among ordinary folk. But one might doubt this view; one might think that when philosophers are trying to figure out what free will and moral responsibility are, they're not just trying to capture ordinary-language meaning. One might think they're also trying to improve upon ordinary usage. But if we ask ourselves what sorts of facts might be relevant here—that is, might be relevant to determining which answers to the what-is-free-will question and the moral responsibility question are correct—there aren't very many plausible candidates. Now, there are certainly a few things one might reasonably say here; for example, one might appeal to facts about the coherence of the various notions of free will and moral responsibility, or perhaps to facts about how well these notions "carve nature at the joints." But I think it can be argued—though, again, I won't argue this point here—that all of the different kinds of facts that one might plausibly appeal to here are either empirical facts or logical facts.

If this is right, and if the various points that I'm going to argue in this book are also right, then it would seem to be a confusion to think of the problem of free will and determinism as a *metaphysical* problem at all—unless by 'metaphysical problem' you simply mean a problem about the nature of reality. If that's all a metaphysical problem is, then the problem of free will is indeed a metaphysical problem, but of course, so are the problems of empirical science. If, however, a metaphysical problem is supposed to be somehow different from the problems that we address in the empirical and logico-mathematical sciences (and if the arguments of this book are cogent), then the problem of free will is not a metaphysical problem, because it reduces to straightforwardly empirical questions and possibly some logical questions.

Now, I actually think the conclusion of the last paragraph can be generalized; I think it can be argued that *all* so-called metaphysical problems reduce to questions that are either empirical (and about the nature of the physical world), or logical, or factually empty in the sense that there are

no facts of the matter about their answers. But again, I will not try to argue for this sweeping conclusion here.

So perhaps the two central aims of this book are to provide (a) a defense of libertarianism and (b) an attack (or at any rate, *part* of an attack) on metaphysics. These two aims might seem like strange bedfellows, because it might seem that libertarianism is the most "metaphysically loaded" view in the free will literature. But by now it should be clear that it is also an aim of this book to "demetaphysicalize" the libertarian view. Once we see (in chapter 3) that libertarianism follows from the relevant kind of indeterminism, we will also see that there is nothing "metaphysically loaded" about libertarianism at all. It's just a straightforward empirical hypothesis about the neural events that are our (torn) decisions. Moreover, if the arguments of chapter 4 are correct, then the question of whether the libertarian hypothesis is true is a *wide open* question. Again, I am going to argue that, at present, we don't have any good reason for taking sides on this question. Thus, I am not going to argue in this book that libertarianism is true. I'm going to defend it against various objections, and I'm going to argue that we do not currently have any good reason to reject it. But I will not argue in its favor, because I also think that we don't have any good reason to endorse it. Again, I think the question of whether libertarianism is true is an open empirical question.