

# Four Views on Free Will

John Martin Fischer,  
Robert Kane,  
Derk Pereboom,  
and  
Manuel Vargas

# Contents

---

<i>Notes on Contributors</i>	vi
<i>Acknowledgments</i>	viii
A Brief Introduction to Some Terms and Concepts	1
1 Libertarianism <i>Robert Kane</i>	5
2 Compatibilism <i>John Martin Fischer</i>	44
3 Hard Incompatibilism <i>Derk Pereboom</i>	85
4 Revisionism <i>Manuel Vargas</i>	126
5 Response to Fischer, Pereboom, and Vargas <i>Robert Kane</i>	166
6 Response to Kane, Pereboom, and Vargas <i>John Martin Fischer</i>	184
7 Response to Kane, Fischer, and Vargas <i>Derk Pereboom</i>	191
8 Response to Kane, Fischer, and Pereboom <i>Manuel Vargas</i>	204
<i>Bibliography</i>	220
<i>Index</i>	224

# Notes on Contributors

---

---

**John Martin Fischer** is Distinguished Professor in the Department of Philosophy at the University of California, Riverside, where he holds a UC President's Chair. He is the author of *The Metaphysics of Free Will: An Essay on Control* (Blackwell, 1994); and, with Mark Ravizza, S.J., *Responsibility and Control: A Theory of Moral Responsibility* (1998). His collection of essays, *My Way: Essays on Moral Responsibility*, was published in 2006.

**Robert Kane** is University Distinguished Teaching Professor of Philosophy at the University of Texas at Austin. He is the author of seven books and over sixty articles on the philosophy of mind and action, ethics, the theory of value and philosophy of religion, including *Free Will and Values* (1985), *Through the Moral Maze: Searching for Absolute Values in a Pluralistic World* (1994), *The Significance of Free Will* (1996), *A Contemporary Introduction to Free Will* (2005), and a lecture series on audio and video tape entitled *The Quest for Meaning: Values, Ethics, and the Modern Experience*. He is editor of *The Oxford Handbook of Free Will* (2002) and of *Free Will* (Blackwell, 2002). The recipient of fifteen major teaching awards at the University of Texas, he was named an inaugural member of the University's Academy of Distinguished Teachers in 1995.

**Derk Pereboom** is Professor of Philosophy at the University of Vermont. He will join the Sage School of Philosophy at Cornell University in 2007. His book *Living Without Free Will* appeared in 2001, and he has published articles on free will, philosophy of mind, history of modern philosophy, and philosophy of religion.

**Manuel Vargas** is Associate Professor of Philosophy at the University of San Francisco. The author of various articles in ethics, philosophy of action, and Latin American philosophy, he has been awarded the American Philosophical Association's Prize in Latin American Thought (2004), and the N.E.H. Chair in the Humanities at the University of San Francisco (2005–2006).

# *A Brief Introduction to Some Terms and Concepts*

---

## **Basic Terms: Free Will, Moral Responsibility, and Determinism**

Perhaps the three most important concepts in philosophical work on free will are *free will*, *moral responsibility*, and *determinism*.

The notion of freedom at stake in philosophical discussions is usually distinguished from a variety of other freedom concepts, including things like religious and political freedom. Usually, **free will** is also treated as distinct from several other concepts associated with human agency, such as autonomy and authenticity. As we will see in the chapters that follow, there are many different ways of thinking about the nature of free will, and there are serious disagreements about what would constitute an adequate theory of free will. Much of the tradition has taken “free will” to be a kind of power or ability to make decisions of the sort for which one can be morally responsible, but philosophers have also sometimes thought that free will might be required for a range of other things, including moral value, originality, and self-governance. Two other claims often made about free will are hotly disputed among philosophers; and authors of this volume will take different sides on these claims. One is the claim that free will requires “alternative possibilities” or the power to do otherwise, and the other is the claim that free will requires that we are the “ultimate sources” of our free actions or the ultimate sources of our wills to perform free actions.

Important to many discussions of free will is the idea of **moral responsibility**. In the context of discussions of free will, moral responsibility is

often understood as a kind of status connected to judgments and/or practices of moral praise and blame. This meaning is distinct from another, perhaps more commonly used sense of responsibility: responsibilities as obligations (for example, when we talk about what responsibilities a parent has to a child). There are important connections between responsibility of the sort concerned with praise and blame and responsibility of the sort connected with obligations. However, philosophers writing on free will and moral responsibility are typically concerned with the former and not the latter.

**Determinism** is a third concept that is often important for philosophical discussions of free will. For present purposes, we can treat determinism as the thesis that at any time (at least right up to the very end) the universe has exactly one physically possible future. Something is deterministic if it has only one physically possible outcome.

It is important to bear in mind that a definition of determinism is just that – a characterization of what things would have to be like *if* things were deterministic. It does not follow that the universe is actually deterministic. Compare: “A creature is a gryphon if it has the hindquarters of a lion and the head and claws of an eagle.” Nothing about the definition of gryphon shows that there are such creatures in our universe. It simply tells us something about what sorts of things would count as gryphons. Similarly, to offer a definition of determinism does not show that the universe is deterministic. It only defines a term, and we may find that the term never properly applies to the world we live in.

When discussing these issues it is natural to wonder whether the world is deterministic. Most physicists and philosophers think that the answer is no, but the technical issues are extremely complex. Nevertheless, if we accept that the universe isn’t deterministic there are still good reasons to think about the compatibility of free will and determinism. First, it could turn out that future physicists conclude that the universe is deterministic, contrary to the contemporary consensus about at least quantum mechanics. It is notoriously difficult to predict how future science will turn out, and it might be useful to have an answer to the question in advance of the scientific issues getting sorted out. Second, even if the universe were not fully deterministic, determinism might hold locally (either as a matter of how local spacetime is constructed, or as a matter of how the physics for non-quantum physical objects operates). Third, we could be interested in whether free will is compatible with a broadly scientific picture of the universe. Since some aspects of the universe seem deterministic and others do not, we might ask if free will is compatible with determinism as a first step to answering the more general question of whether free will is compatible with a broadly scientific picture of the universe.

## Philosophical Options on the Free Will Problem

One particularly important issue for contemporary philosophers thinking about free will is whether we could have free will in a deterministic universe. Call this issue – whether free will could exist if the universe were deterministic – **the compatibility issue**. There is a long-standing tradition of dividing up the conceptual terrain in light of the main answers to the compatibility issue. Traditionally, **incompatibilists** are those who think that free will is incompatible with the world being deterministic. **Compatibilists**, conveniently enough, are those hold that free will is compatible with the universe being deterministic.

It is important to recognize that the compatibility issue is distinct from the issue of whether we have free will. You could be an incompatibilist, and maintain that we have do have free will. Or you might be an incompatibilist and think that we lack free will. (You could even think that irrespective of how the compatibility issue is settled, there are threats to free will apart from determinism.)

In the philosophical literature, **libertarianism** is the view that we have free will and that free will is incompatible with determinism. “Libertarianism” as it is used in the context of free will is distinct from libertarianism in political philosophy. (Indeed, “libertarianism” in the free will sense is the original meaning – it was only later appropriated as the label for a view in political philosophy.) One might be a libertarian in both political and free will senses, but you can be a libertarian about free will without being a libertarian in political philosophy. And, perhaps, you could also be a political libertarian without being a free will libertarian (although many political libertarians seem to also be free will libertarians).

Following Derk Pereboom, we will label as “**hard incompatibilism**” any view that holds that (1) incompatibilism is true and (2) we lack free will. Historically, most hard incompatibilists were what William James called **hard determinists**. (Indeed, Pereboom’s coining of the term “hard incompatibilism” reflects James’ older and narrower terminology.) Hard determinists think we lack free will *because the world is deterministic*. Contemporary hard determinists are few and far between. What is more common are views that hold that we have no free will irrespective of whether or not the world is deterministic, and views that hold that although freedom might be not be conceptually incompatible with determinism (or indeterminism, for that matter), we simply do not have it.

To summarize, then: A traditional way of dividing up the terrain concerns answers to the compatibility issue. The two main approaches are incompatibilism and compatibilism. We have been considering the incompatibilist fork, where the two main species of incompatibilism are libertarianism and hard

incompatibilism. Both forms of incompatibilism have further species we have not discussed in this brief introduction.

The remaining fork of the compatibility debate is **compatibilism**. There are many varieties of compatibilism. Some compatibilists have emphasized a particular understanding of “can,” others have emphasized a kind of identification with one’s motives or values, and others emphasizing the role of responsiveness to reasons. One influential variation, however, is the view that holds that responsibility is compatible with determinism, combined with agnosticism about whether free will understood in some particular way might not be compatible with determinism. This view is *semicompatibilism*, and its most prominent defender is John Martin Fischer.

Lastly, there are views that do not neatly fit the traditional taxonomy of incompatibilism and compatibilism. One such class of views is **revisionism**. The core idea of revisionism is that the picture of free will and moral responsibility embedded in commonsense is in need of revision, but not abandonment. That is, the revisionist holds that the correct account of free will and moral responsibility will depart from commonsense. As is the case with libertarianism, hard incompatibilism, and compatibilism, this view can take a variety of more specific forms.

For a different way to think about the relationship between the various views, see the grid below.

	Is commonsense thinking about free will and moral responsibility basically correct?	Is free will compatible with determinism?	Is moral responsibility compatible with determinism?	Do we have free will?
Libertarianism	Yes	No	No	Yes
Compatibilism	Yes	Yes (although semicompatibilists may say “no”)	Yes	Yes
Hard Incompatibilism	No	No	No	No
Revisionism	No	Yes, but only with revision to our self-image	Yes	Yes (or close enough)



# 1

## *Libertarianism*

---

---

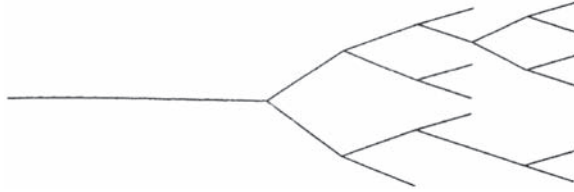
*Robert Kane*

### **1 Determinism and the Garden of Forking Paths**

The problem of free will has arisen in history whenever people have been led to suspect that their actions might be determined or necessitated by factors unknown to them and beyond their control. That is why doctrines of *determinism* or *necessity* have been so important in the history of debates about free will.

Doctrines of determinism have taken many historical forms. People have wondered at various times whether their actions might be determined by Fate or by God, by the laws of physics or the laws of logic, by heredity or environment, by unconscious motives or hidden controllers, psychological or social conditioning, and so on. But there is a core idea running through all historical doctrines of determinism that shows why they are all a threat to free will. All doctrines of determinism – whether they are fatalistic, theological, physical, biological, psychological or social – imply that, given the past and the laws of nature at any given time, there is only one possible future. Whatever happens is therefore inevitable or necessary (it cannot but occur), given the past and the laws.

To see why many persons have believed there is a conflict between free will and determinism, so conceived, consider what free will requires. We believe we have free will when we view ourselves as agents capable of influencing the world in various ways. Open alternatives seem to lie before us. We reason and deliberate among them and choose. We feel (1) it is “up to us” what we choose and how we act; and this means we could have chosen or acted otherwise. As Aristotle said, “when acting is ‘up to us,’ so is not acting.” This “up-to-us-ness” also suggests that (2) the ultimate sources of our actions lie in us and not outside us in factors beyond our control.



**Figure 1** Garden of Forking Paths

To illustrate, suppose Jane has just graduated from law school and she has a choice between joining a law firm in Chicago or a different firm in New York. If Jane believes her choice is a *free* choice (made “of her own free will”), she must believe both options are “open” to her while she is deliberating. She could choose either one. (If she did not believe this, what would be the point of deliberating?) But that means she believes there is more than one possible path into the future available to her and it is “up to her” which of these paths will be taken. Such a picture of an open future with forking paths – a garden of forking paths, it has been called – is essential to our understanding of free will.

This picture of different possible paths into the future is also essential, I believe, to what it means to be a person and to live a human life.

One can see why determinism would threaten this picture. If determinism is true, it seems there would not be more than one possible path into the future available to Jane, but only one. It would not be (1) “up to” her what she chose from an array of alternative possibilities, since only one alternative would be possible. It also seems that, if determinism were true, the (2) sources or origins of her actions would not be in Jane herself but in something else outside her control that determined her choice (such as the decrees of fate, the foreordaining acts of God, her heredity and upbringing or social conditioning).

A second way to illustrate why many people believe there is a conflict between free will and determinism is to reflect on the idea of *responsibility*. Free will is also intimately related to notions of accountability, blameworthiness and praiseworthiness for actions.

Suppose a young man is on trial for an assault and robbery in which his victim was beaten to death. Let us say we attend his trial and listen to the evidence in the courtroom. At first, our thoughts of the young man are filled with anger and resentment. His crime was heinous. But as we listen daily to how he came to have the mean character and perverse motives he did have – a sad story of parental neglect, child abuse, sexual abuse, bad role models –

some of our resentment against the young man is shifted over to the parents and others who abused and mistreated him. We begin to feel angry with them as well as with him. (Note how natural this reaction is.) Yet we aren't quite ready to shift all of the blame away from the young man himself. We wonder whether some residual responsibility may not belong to him. Our questions become: To what extent is *he* responsible for becoming the sort of person he now is? Was it *all* a question of bad parenting, societal neglect, social conditioning, and the like, or did he have any role to play in it?

These are crucial questions about free will and they are questions about what may be called the young man's ultimate responsibility. We know that parenting and society, genetic make-up and upbringing, have an influence on what we become and what we are. But were these influences entirely *determining* or did they "leave anything over" for us to be responsible for? That is what we want to know about the young man. The question of whether he is merely a victim of bad circumstances or has some responsibility for being what he is – the question, that is, of whether he became the person he is *of his own free will* – seems to depend on whether these other factors were or were not *entirely* determining.

Those who are convinced that there is a conflict between free will and determinism, for these and other reasons, are called *incompatibilists* about free will. They believe free will and determinism are incompatible. If incompatibilists also believe that an incompatibilist free will exists, so that determinism is false, they are called *libertarians* about free will.

## 2 Modern Challenges to Libertarian Free Will

I will be defending the libertarian view of free will in this volume. We libertarians typically believe that a free will that is incompatible with determinism is required for us to be truly morally responsible for our actions, so that genuine moral responsibility, as well as free will, is incompatible with determinism. Genuine free will, we believe, could not exist in a world that was *completely* determined by Fate or God, or the laws of physics or logic, or heredity and environment, psychological or social conditioning, and so on. In writings over the past twenty-five years, I have argued that this libertarian view represents the traditional idea of free will that has been in dispute for centuries when philosophers have discussed "the problem of free will and determinism." Moreover, I think this libertarian view is the one many ordinary persons have in mind when they intuitively believe there is some kind of conflict between free will and determinism.

Yet this traditional libertarian conception of free will has been under attack by many modern thinkers, philosophers and scientists alike, who have come

to believe that such an idea of free will, though it may still be held by many ordinary people, is outmoded and incoherent and that it has no place in the modern scientific picture of the world. A goal of this essay is therefore to consider this modern attack on the traditional libertarian view of free will and to ask how, and whether, it can be answered. Much is at stake, it seems to me, in knowing whether we do or do not have a freedom of the will of the ultimate kind that libertarians defend. The modern attack on it has two parts.

*Part 1* The first prong of the modern attack on libertarian free will comes from *compatibilists*, who argue that, despite appearances to the contrary, determinism does not really conflict with free will at all. Compatibilists argue that all the freedoms we recognize and desire in ordinary life – e.g., freedoms from coercion or compulsion, from physical restraint, from addictions and political oppression, for example – are really compatible with determinism. Even if the world should turn out to be entirely deterministic, compatibilists argue, there would still be a big difference between persons who are free from constraints on their freedom of action and will (constraints such as coercion, compulsion, addiction and oppression) and persons who are not free from these constraints; and people would prefer to be free from such constraints on their freedom rather than not, *even in a determined world*. Thus, according to compatibilists, esoteric questions about whether determinism is true or not – in the physical or psychological sciences – are irrelevant to *the freedoms we really care about* in everyday life. All the varieties of free will “worth wanting” (as a modern compatibilist, Daniel Dennett, has put it) do not require the falsity of determinism for us to possess them, as the traditional libertarian view of free will suggests.

This doctrine of *compatibilism* has an ancient lineage. It was held by the Stoics and perhaps also by Aristotle in ancient times, according to many scholars. But compatibilism about free will and determinism has become especially popular in modern times. Influential philosophers of the modern era, such as Thomas Hobbes, John Locke, David Hume and John Stuart Mill, were all compatibilists. They saw compatibilism as a way of reconciling ordinary experience of being free with modern scientific views about the universe and human beings; and compatibilism continues to be popular among philosophers and scientists today for similar reasons, as you will see from later essays of this volume. (John Martin Fischer defends a version of compatibilism, known as *semicompatibilism*, in the second essay of this volume.) If compatibilists are right, we can have *both* free will and determinism; and we need not worry that increasing scientific knowledge about nature and human beings will somehow undermine our ordinary convictions that we are free and responsible agents.

*Part 2* The second prong of the modern attack on libertarian free will goes a step further. Recall that the first prong says that libertarian free will is *unnecessary* because we can have all the freedoms worth wanting, even if determinism should be true. The second prong goes further, arguing that libertarian free will itself is *impossible* or *unintelligible* and has no place in the modern scientific picture of the world. Such an ultimate freedom is not something we could have anyway, say its critics. Those who take this line note that defenders of libertarian free will have often invoked obscure and mysterious forms of agency or causation to defend the libertarian view. In order to explain how free actions can escape the clutches of physical causes and laws of nature (so that free actions will not be determined by physical laws), libertarians have posited transempirical power centers, immaterial egos, noumenal selves outside of space and time, unmoved movers, uncaused causes and other unusual forms of agency or causation – thereby inviting charges of obscurity or mystery against their view. Even some of the greatest modern defenders of libertarianism, such as Immanuel Kant, have argued that we need to believe in libertarian free will to make sense of morality and genuine responsibility, but we can never completely understand such a free will in theoretical and scientific terms.

The problem that provokes this widespread skepticism about the existence of libertarian free will has to do with an ancient dilemma: If free will is not compatible with *determinism*, as libertarians contend, free will does not seem to be compatible with *indeterminism* either (the opposite of determinism). Events that are undetermined, such as quantum jumps in atoms, happen merely by chance. So if free actions were undetermined, as libertarians claim, it seems that they too would happen by chance. But how can chance events be free and responsible actions? Suppose a choice was the result of a quantum jump or other undetermined event in a person's brain. Would this amount to a free and responsible choice? Undetermined effects in the brain or body would be unpredictable and impulsive – like the sudden emergence of a thought or the uncontrolled jerking of an arm – quite the opposite of what we take free and responsible actions to be. It seems that undetermined events in the brain or body would occur *spontaneously* and would be more likely to *undermine* our freedom rather than *enhance* it.

This two-pronged modern attack on the traditional libertarian view of free will has had a powerful impact on modern thought. To answer it, libertarians must show (i) that free will really *is* incompatible with *determinism* (call this “The Compatibility Problem”). But they must also show (ii) that a libertarian free will requiring *indeterminism* can be made intelligible and how, if at all, such a free will can be reconciled with modern scientific views of the cosmos and of human beings (call this “The Intelligibility Problem”). I will be

addressing both these problems in this chapter, beginning with the first, or “Compatibility Problem.”

### 3 Is Free Will Incompatible with Determinism?: The Consequence Argument

The popularity of compatibilism among modern philosophers and scientists means that libertarians who believe free will is incompatible with determinism can no longer merely rely on intuitions about “forking paths” into the future to support their view that determinism conflicts with free will (as in section 1). These intuitions must be backed up with arguments that show *why* free will must be incompatible with determinism. To meet this challenge, libertarians have proposed new arguments for incompatibilism in modern philosophy; and we will begin by considering the most widely discussed of these new arguments for the incompatibility of free will and determinism.

This important argument is called the “Consequence Argument” and it is stated informally as follows by one of its proponents, Peter van Inwagen:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our own acts) are not up to us. (From *An Essay on Free Will*, Oxford: Clarendon Press, 1983, p. 16)

To say it is not “up to us” what “went on before we were born,” or “what the laws of nature are,” is to say that there is nothing we can now do to change the past or alter the laws of nature (it is beyond our control). We can thus spell out this Consequence Argument in the following steps:

- (1) There is nothing we can now do to change the past.
- (2) There is nothing we can now do to change the laws of nature.
- (3) There is nothing we can now do to change the past and the laws of nature.
- (4) If determinism is true, our present actions are necessary consequences of the past and the laws of nature. (That is, it *must* be the case that, given the past and the laws of nature, our present actions occur.)
- (5) Therefore, there is nothing we can now do to change the fact that our present actions occur.

In other words, we *cannot now do otherwise* than we actually do. Since this argument can be applied to any agents and actions at any times, we can infer

from it that *if determinism is true, no one can ever do otherwise*; and if free will requires the power to do otherwise than we actually do (as in the image of forking paths), then no one would have free will.

Defenders of the Consequence Argument, such as van Inwagen, think the first two premises are undeniable. We cannot now change the past (1) or the laws of nature (2). Step 3 states what appears to be a simple consequence of premises 1 and 2: If you can't change the past or the laws, then you can't change the conjunction of both of them. Premise 4 simply spells out what is implied by determinism. Some philosophers have questioned one or another of the first three steps of this argument. But most criticisms have focused on step 5. Step 5 follows from 3 and 4 by virtue of the following inference: If (3) there is nothing we can now do to change the past and laws of nature and (4) our present actions are necessary consequences of the past and laws, then (5) there is nothing we can now do to change the fact that our present actions occur. This inference is an instance of the following principle:

(TP) If there is nothing anyone can do to change X, and if Y is a necessary consequence of X (if it must be that, if X occurs, Y occurs), then there is nothing anyone can do to change Y.

TP has been called a "Transfer of Powerlessness Principle" for it says in effect that if you are powerless to change something X, and something else Y is necessarily going to occur if X does, then you are also powerless to change Y. This makes sense. If we can't do anything to prevent X from occurring and Y cannot but occur if X does, then how could we do anything to prevent Y from occurring? Consider an example. Suppose the sun is going to explode in AD 2050 and there is nothing anyone can now do to change the fact that the sun will explode in AD 2050. Assume also that necessarily (given the laws of nature), if the sun explodes in AD 2050, all life on earth will end in AD 2050. If both these claims are true, it seems obvious that there is nothing anyone can now do to change the fact that all life on earth will end in 2050. Here is another example. If there is nothing anyone can now do to change the laws of nature, and the laws of nature entail that nothing goes faster than the speed of light, then there is nothing anyone can now do to change the fact that nothing goes faster than the speed of light.

But, despite the initial plausibility of this Transfer of Powerlessness Principle, critics of the Consequence Argument have challenged it. Everything depends, they say, on how you interpret the expression "There is nothing anyone can do to change . . ." Talking about what persons "can" (and "cannot") do is talking about their *powers*; and the notion of power is one of the most difficult in metaphysics, as John Locke pointed out three centuries ago. For example, many *compatibilists* interpret what it means to

say that persons “can” or “have the power” to do things in the following way. They say

“You can (or you have the power to do) something.”

simply means

“If you wanted (or tried) to do it, you would do it.”

I can jump over this fence means I would jump over it, if I wanted to or tried to. If someone challenged my power to do it, the challenger would say “I don’t think you would manage to jump it, *even if* you wanted or tried.”

Now the interesting thing about this compatibilist interpretation of “can” and “power,” is that, if it is correct, the Consequence Argument would fail. For on this interpretation, to say we can now change the past or the laws would mean that

“*If* we now wanted or tried to change the past or the laws, we would change them.”

And this is false. No persons would change the past or the laws of nature, *even if* they wanted or tried to, because no one has the power to do it. But when we turn to ordinary actions like jumping over a fence, things are different. If you can jump over a fence that is in your path, it may well be true that you *would* jump over it, *if* you wanted to or tried, because jumping over fences is something you *are* capable of doing.

In other words, on the analysis of “can” or “power” that many compatibilists favor, the *premises* of the Consequence Argument come out *true* (you would *not* have changed the past or the laws, even if we wanted or tried to, because you are not capable of it). But the *conclusion* of the Consequence Argument comes out *false* (you would have jumped the fence, *if* you wanted or tried to, because jumping fences of this height is something you *are* capable of doing). Since the Consequence Argument would have true premises and a false conclusion on this analysis of “can,” it would be an invalid argument. What has happened to make it fail? The answer is that the transfer principle TP has failed. Your powerlessness to change the past and laws of nature does not *transfer* to your powerlessness to jump the fence. For you are *not* able to change the past and laws, but you *are* able to jump the fence – at least in this compatibilist sense that (“you would do it, *if* you wanted or tried to.”

But why should we accept this “hypothetical” compatibilist account of “can” or “power” (“you would do it, *if* you wanted or tried to”)? Defenders of the Consequence Argument, such as van Inwagen, do not accept this



hypothetical account of “can” or “power”; nor do most libertarians. They would respond to the preceding compatibilist argument as follows:

“So the Consequence Argument fails on your compatibilist analysis of ‘can’ or ‘power.’ But that should not surprise us. For your compatibilist analysis was rigged in the first place to make freedom compatible with determinism. On your analysis, persons can jump the fence even though their doing so here and now is impossible, given the past and the laws of nature. That is not what we libertarians mean by ‘can’ in the Consequence Argument. We mean it is possible that you do it *here and now, given all the facts that presently obtain*. If your analysis allows you to say that persons can do otherwise, even though they can’t change the past and the laws of nature and even though their actions are a necessary consequence of the past and the laws of nature, *then something must be wrong with your compatibilist analysis*. What use is a power or ability to do something, if it cannot be *exercised* in the existing circumstances here and now? To us libertarians, the premises and rules of the Consequence Argument are far more plausible than any compatibilist analysis of ‘can.’”

At this point, arguments over the Consequence Argument tend to reach an impasse. Incompatibilist defenders of the argument claim that compatibilist critics are begging the question by interpreting “can” in the Consequence Argument in a way that is compatible with determinism. But compatibilists respond by saying that defenders of the Consequence Argument are begging the question themselves by assuming that “can” in the argument has an *incompatibilist* meaning rather than a compatibilist one.

#### 4 Ultimate Responsibility

As a result of this impasse, philosophical debates have multiplied about just what “can” and “power” (and related expressions, such as “could have done otherwise”) really mean. We cannot follow all these complex debates here. But I do not think it matters. For I believe disagreements over the meaning of “can” and “power” are symptoms of a deeper problem in discussions about free will and determinism. The problem is that focusing on “alternative possibilities” (or “forking paths” into the future) or the “power to do otherwise” *alone*, as the Consequence Argument does, is *too thin a basis* on which to rest the case for the incompatibility of free will and determinism. One must look beyond debates about “can,” “power,” “ability,” and “could have done otherwise” to make the case for the incompatibility of free will and determinism.

Fortunately, there is another place to look for reasons why free will might conflict with determinism. Recall that in section 1, I suggested that there were *two* reasons why people thought determinism must rule out free will.

One was the requirement of (1) alternative possibilities we have been considering: Free will seems to require that *open alternatives* or *alternative possibilities* lie before us – a garden of forking paths – and it is “up to us” which of these alternatives we choose. (Call this condition “AP” for “alternative possibilities”). But there was a second condition mentioned that has also historically fueled incompatibilist intuitions: (2) Free will also seems to require that the *sources* or *origins* of our actions lie “in us” rather than in something else (such as the decrees of fate, the foreordaining acts of God, or antecedent causes and laws of nature) outside us and beyond our control.

I call this second requirement for free will the condition of Ultimate Responsibility (or UR, for short); and I think it is even more important to free will debates than AP, or alternative possibilities. The basic idea of UR is this: *To be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient cause or motive for the action's occurring.* If, for example, a choice issues from, and can be sufficiently explained by, an agent's character and motives (together with background conditions), then to be *ultimately* responsible for the choice, the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past for having the character and motives he or she now has. Compare Aristotle's claim that if a man is responsible for the wicked acts that flow from his character, he must at some time in the past have been responsible for forming the wicked character from which these acts flow.

This condition of Ultimate Responsibility, or UR, does not require that we could have done otherwise (AP) for *every* act done of our own free wills. But it does require that we could have done otherwise with respect to *some* acts in our past life histories by which we formed our present characters. I call these earlier acts by which we formed our present characters “self-forming actions,” or SFAs.

To see why such self-forming acts are important for free will, consider a well-known example about Martin Luther offered by Daniel Dennett. When Martin Luther finally broke with the Church in Rome, initiating the Protestant Reformation, he said “Here I stand, I can do no other.” Now Dennett asks us to suppose that at the moment Luther made this stand, he was literally right. Given his character and motives, Luther *could* not then and there *have done otherwise*. Does this mean Luther was not morally responsible, not subject to praise or blame, for his act, or that he was not acting of his own free will? Dennett says “not at all.” In saying “I can do no other,” Luther was not disowning responsibility for his act, according to Dennett, but taking full responsibility for acting of his own free will. So the ability to do otherwise (“could have done otherwise”) or AP, says Dennett, is not required for moral responsibility or free will.

Now Dennett is a compatibilist, as noted earlier, and he is using this Luther example to defend compatibilism of free will and determinism by

suggesting that free will and moral responsibility do not even require the power to do otherwise or alternative possibilities (AP). Note that, if this were true, the Consequence Argument would be undermined. We would not have to get into complex debates about what “could have done otherwise” means, since free will and moral responsibility would not require alternative possibilities (AP) or “could have done otherwise” in the first place.

But, now, if we look at Dennett’s Luther example from the point of view of the condition of Ultimate Responsibility or UR, rather than simply in terms of AP, there is an answer that can be given to Dennett. We can grant that Luther could have been responsible for this act, even though he could *not* have done otherwise *then and there* and even if his act was determined. But this would be so, if UR is required, only to the extent that Luther was responsible for his present motives and character by virtue of some *earlier* struggles and self-forming actions (SFAs) that brought him to this point in his life where he could do no other. Those who know Luther’s biography know the inner struggles and turmoil he endured getting to that point in his life. Often we act from a will already formed, but it is “our own free will” by virtue of the fact that *we* formed it by other choices or actions in the past (SFAs) for which we *could* have done otherwise. If this were not so, *there is nothing we could have ever done to make ourselves different than we are* – a consequence, I believe, that is incompatible with our being (at least to some degree) ultimately responsible (UR) for what we are. So SFAs are only a subset of those acts in life for which we are ultimately responsible and which are done “of our own free will.” But if *none* of the acts in our lifetimes were self-forming in this way, we would not be *ultimately* responsible for anything we did.

If the case for incompatibility of free will and determinism cannot be made by reference to AP alone, it can be made if UR is added. So, I suggest, the often-neglected condition of ultimate responsibility or UR should be moved to center stage in free will debates. If agents must be responsible to some degree for anything that is a *sufficient cause* or *motive* for their actions (as UR requires), then an impossible infinite regress of past actions would be required unless some actions in the agent’s life history (SFAs) did not have either sufficient causes or motives (and hence were undetermined). Therein lies the connection between UR and determinism. If we must have formed our present wills (our characters and motives) by earlier voluntary choices or actions, then UR would require that if any of these earlier choices or actions *also* had sufficient causes or motives when we performed *them*, then we must have also been responsible for those earlier sufficient causes or motives by virtue of forming them by *still earlier* voluntary choices or actions, and so on backwards indefinitely into our past. Eventually we would come to infancy or to a time before our birth when we could not have formed our own wills.

The only way to stop this regress is to suppose that *some* acts in our life histories must lack *sufficient* causes altogether, and hence must be undetermined, if we are to be the ultimate sources or grounds of, and hence ultimately responsible for, our own wills. These regress-stopping acts would be the “self-forming acts” or SFAs that are required by UR sometime in our lives, if we are to have free will. Note, as a result, that UR makes explicit something that is often hidden in free will debates, namely that *free will* – as opposed to mere *freedom of action* – is about the forming and shaping of character and motives which are the *sources* or *origins* of praiseworthy or blameworthy, virtuous or vicious, actions. *Free will* (in contrast to mere *free action*) is about *self-formation*. If persons are responsible for the wicked (or noble, shameful, heroic, generous, treacherous, kind or cruel) acts that flow from their wills (characters and motives), they must at some point be responsible for forming the *wills* from which these acts flow.

## 5 Ultimate Responsibility and Alternative Possibilities

Another thing to note about this argument for the incompatibility of free will and determinism from UR is that – unlike the Consequence Argument – the argument from UR does not mention the condition of *alternative possibilities* or AP at all. The argument from UR says that, if agents must be responsible to some degree for anything that is a *sufficient cause or motive* for their actions (as UR requires), then an impossible infinite regress of past actions would be required, *unless* some actions in the agent’s life history (SFAs) did not have either sufficient causes or motives and hence were undetermined. The argument from UR thus focuses on the sources or origins of what we actually do rather than on the power to do otherwise.

When one argues about the incompatibility of free will and determinism from alternative possibilities or AP (as in the Consequence Argument), the focus is on notions of “necessity,” “possibility,” “power,” “ability,” “can,” and “could have done otherwise.” By contrast, the argument from UR focuses on a different set of concerns about the “sources,” “grounds,” “reasons,” and “explanations” of our wills, characters, and purposes. Where did our motives and purposes come from, who produced them, who is responsible for them? Was it *we* ourselves who are responsible for forming our characters and purposes, or someone or something else – God, fate, heredity and environment, nature or upbringing, society or culture, behavioral engineers or hidden controllers? Therein lies the core of the traditional problem of free will.

But does this mean that alternative possibilities or AP have nothing to do with free will? It might seem so, if one can argue directly for the incompatibility of free will and determinism from UR without mentioning alternative

possibilities. But then what would become of the garden of “forking paths” if alternative possibilities or AP are not required? Well, fortunately it turns out that AP and the garden of forking paths *is* relevant for free will after all. For, it can be shown that *UR entails AP* for at least *some* free actions. Why this is so is not obvious, but understanding it is also crucial, I believe, to fully understand the nature of free will.

To understand the connection between AP and UR, alternative possibilities and ultimate responsibility, we must first note that having alternative possibilities for one’s action – though it may be necessary for free will – is not *sufficient* for free will, *even if* the alternative possibilities should also be *undetermined*. This can be shown by noting that there are examples in which agents may have alternative possibilities *and* their actions are undetermined, *and yet the agents lack free will*. This sounds strange. But it is important for understanding free will to understand how it could be. I call examples that show this “Austin-style examples” after the British philosopher J. L. Austin, who suggested the first example of this kind in free will debates.

Here are three easily understood “Austin-style examples” that I will refer to in later arguments. The first example is Austin’s own. (i) He imagined that he had to hole a three-foot putt to win a golf match but, owing to a nervous twitch in his arm, he misses the putt. The other two examples are mine. (ii) An assassin is trying to kill the prime minister with a high-powered rifle when, owing to a nervous twitch in his arm, he misses and kills the minister’s aide instead. (iii) I am standing in front of a coffee machine intending to press the button for coffee without cream when, owing to a brain cross, I accidentally press the button for coffee with cream. Now notice that in each of these examples, we can suppose, as Austin suggests, that an element of genuine chance or indeterminism is involved. Perhaps the nervous twitches or brain crosses are brought about by actual undetermined quantum jumps in our nervous systems. We can thus imagine that Austin’s holing the putt is a genuinely undetermined event. He might miss the putt by chance and, in the example, does miss it by chance. (Likewise, the assassin might hit the wrong target by chance and I might press the wrong button by chance.)

Now Austin asked the following question about his example: Can we say in these circumstances that “he (Austin) could have done otherwise” than miss the putt? Did he have alternative possibilities? Austin’s answer is that we can indeed say he could have done otherwise than miss it. For he was a good putter. He had made many similar putts of this short length in the past (he had the capacity and opportunity to make it). But even more important, since the outcome of this putt was genuinely *undetermined*, he might well have succeeded in holing the putt, as he was trying to do.

But this means we have an action (missing the putt) that is (i) *undetermined* and (ii) such that the agent could have done otherwise. (In other words, we

have indeterminism *plus* alternative possibilities or AP.) Yet missing the putt is not something that we regard as *freely* done in any normal sense of the term because it is not under the agent's voluntary control. Austin missed the putt all right; and he *could* have holed it – he could have done otherwise. But he did not miss it *voluntarily* and *freely*. He did not choose to miss it. The same is true of the assassin's failing to hit the prime minister and killing the aide and my accidentally pressing the wrong button on the coffee machine. Both of us could have done otherwise (the assassin could have hit his target and I could have pressed the right button) because our actions were undetermined and they might have gone the other way. Yet the assassin did not miss his target voluntarily and as a result of his own free choice; and I did not press the wrong button voluntarily and as a result of my own free choice.

One might be tempted to think that these three occurrences (missing the putt, killing the aide, pressing the wrong button) are not *actions* at all in such circumstances because they are undetermined and happen by accident. But Austin correctly warns against drawing such a conclusion. Missing the putt, he says, was clearly something he *did*, even though it was not what he *wanted* or *chose* to do. Similarly, killing the aide was something the assassin did, though unintentionally; and pressing the wrong button was something I did, even if only by accident or inadvertently. Austin's point is that many of the things we do *by accident* or *mistake*, *unintentionally* or *inadvertently*, are nonetheless things we *do*. We may sometimes be absolved of responsibility for doing them (though not always, as in the case of the assassin). But it is for *doing* them that we are absolved of responsibility; and this can be true even if the accidents or mistakes are genuinely undetermined.

But now we can draw a further conclusion from these Austin-style examples (the conclusion we were looking for) that Austin himself did not consider. These examples also show that alternative possibilities *plus* indeterminism are not sufficient for *free will* (even if they should be necessary). To see why, suppose that God created a world in which there is a lot of indeterminism of the kind that occurs in Austin-style examples. Chance plays a significant role in this world, in human affairs as well as in nature. People set out to do things and often succeed, but sometimes they fail in the manner of Austin-style examples. They set out to kill prime ministers, hole putts, press buttons on coffee machines, thread needles, punch computer keys, scale walls, and so on – usually succeeding, but sometimes failing by mistake or accident in ways that are undetermined.

Now imagine further that in this world all actions of all agents, whether they succeed in their purposes or not, are such that their reasons, motives and purposes for trying to act as they do are always predetermined or pre-set by God. Whether the assassin misses the prime minister or not, his intent to kill the prime minister in the first place is predetermined by God. Whether

or not Austin misses his putt, his wanting and trying to make it rather than miss are preordained by God. Whether I press the button for coffee without cream, my wanting to do so because of my dislike of cream is predetermined by God; and so it is for all persons and all of their actions in this imagined world. Their reasons, motives and purposes for acting as they do are always predetermined by God.

I would argue that persons in such a world lack *free will*, even though it is often the case that they can do otherwise (and thus have alternative possibilities) in a way that is undetermined. The reason is that they can do otherwise, but only in the limited Austin-style way – by mistake or accident, unwillingly or unintentionally. What they cannot do in any sense is *will* otherwise than they do; for all of their reasons, motives and purposes have been pre-set by God. We may say that the wills of persons in this world are always already “set one way” before and when they act, so that if they do otherwise, it will not be “in accordance with their wills.”

The possibility of such worlds shows in a striking way why, to have free will, it is necessary not only to be the ultimate source of one’s *actions*, but also to be the ultimate source of one’s *will* to perform the actions. It would not be enough for free will to be unhindered in the pursuit of one’s motives and purposes, if all of one’s motives and purposes were created by someone or something else (God or fate or whatever) as in the above-imagined world. Even one’s motives or purposes for wanting to change one’s motives or purposes would be created by someone or something else in such a world.

Now UR captures this additional requirement of being the ultimate source of one’s *will* that is lacking in this imagined world. For UR says that we must be responsible by virtue of our voluntary actions for anything that is a sufficient cause or a sufficient *motive* (or *reason*) for our acting as we do. We have a sufficient motive or reason for doing something, when our will is “set one way” on doing it before and when we act – as the assassin’s will is set on killing the prime minister. Among the available things he might do, only one of them (killing the prime minister) would be voluntary and intentional. Anything else he might do (miss the prime minister, kill the aide) would be done only by accident or mistake, unintentionally or unwillingly.

UR says that if you have a sufficient motive for doing something in this sense – if your will is “set one way” on doing it rather than anything else available to you – then to be ultimately responsible for your *will*, you must be to some degree responsible by virtue of past voluntary acts for your will’s being set the way it is. This is significant because, when we look to the responsibility of the assassin for what he did, we look to his evil motives and intentions. They are the source of his guilt, whether he succeeds in killing the prime minister or fails and kills the aide instead. Luther too, we assumed,

had a sufficient motive for his final affirmation, “Here I stand.” Yet, we said that if Luther’s will was firmly set one way by the time he made his affirmation, this would not count against his being ultimately responsible, *so long as he was responsible for his will’s being set that way*. That is what UR requires.

But now it looks like we have another regress on our hands. If it should turn out that our wills were already set one way when we performed the earlier voluntary actions *by* which we set our present wills, then UR would require that we must have been responsible by virtue of still earlier voluntary actions for our wills’ being set the way they were at that earlier time, and so on backwards indefinitely. But, once again, this is only a *potential* regress. Just as the regress discussed earlier could be stopped by assuming that some actions in an agent’s history lacked *sufficient causes*, so this regress can be stopped by supposing that some actions in an agent’s past also lacked *sufficient motives*. Actions lacking sufficient motives would be actions in which the agents’ wills were not already set one way *before* they performed them. Rather, the agents would set their wills one way or another in the performance of the actions themselves.

We may call such actions in which agents “set their wills” in one way or another in the performance of the actions themselves “will-setting” actions. Will-setting actions occur, for example, when agents make choices or decisions between two or more competing options and do not settle on which of the options they want more, all things considered, until the moment of choice or decision itself. They thus “set” their wills in one way or the other in the act of choosing itself.

The need for such will-setting actions tells us something further about free will. When we wonder about whether agents have freedom of will (rather than wondering only about whether they have freedom of action), what interests us is not merely whether they could have done otherwise, even if the doing otherwise is undetermined, but whether they could have done otherwise *voluntarily* (or *willingly*), *intentionally*, and *rationally*. Or, more generally, we are interested in whether they could have acted in *more than one way* voluntarily, intentionally, and rationally, rather than (as in the Austin-style examples) only in one way voluntarily, intentionally, and rationally and in other ways merely by accident or mistake, unintentionally or irrationally. (“Voluntarily” means here “in accordance with one’s will”; “intentionally” means “knowingly” and “on purpose” and “rationally,” means “having good reasons for acting and acting for those reasons.”)

We might call these requirements of *more-than-one-way* (or plural) voluntariness, rationality, and intentionality, “*plurality conditions*” for free will. Such conditions seem to be deeply embedded in our intuitions about free choice and action. Most of us naturally assume that freedom and responsibility would be deficient if it were always the case that we could only do otherwise



by accident or mistake, unintentionally, or involuntarily. Free will seems to require that if we acted voluntarily, intentionally, and rationally, we could also have done otherwise voluntarily, intentionally, and rationally. But *why* do we assume this so readily; and why are these plurality conditions so deeply embedded in our intuitions about free will?

The argument of the previous section from UR provides the clue. If (i) *free will* requires (ii) *ultimate responsibility* for our wills as well as for our actions, then it requires (iii) *will-setting* actions at some points in our lives; and will-setting actions require (iv) the *plurality conditions*, the ability to act in more than one way voluntarily, intentionally and rationally. To see why will-setting actions require the plurality conditions, consider a variation on the assassin example that would make his choice to kill the prime minister a will-setting one. Suppose that just before pulling the trigger, the assassin has doubts about his mission. Pangs of conscience arise in him and a genuine inner struggle ensues about whether or not to go through with the killing. The assassin now has more than one motivationally significant option before his mind. So his will is no longer clearly set one way; and he will only resolve the issue one way or the other by consciously deciding and thereby setting his will in one direction or the other. Unlike the original assassin example, neither outcome in this case (where he is conscience-stricken and has to decide one way or the other) would be a mere accident or mistake; either resolution would be a voluntary and intentional decision to go through with the killing or to stop. Such a will-setting action would therefore be voluntary, intentional, and rational whichever way it goes and so it would satisfy the plurality conditions.

So we have the following chain of inferences: (1) *free will* entails (2) *ultimate responsibility* [UR] for our wills as well as for our actions, which entails (3) *will-setting* actions at some points in our lives, which in turn entails that some of our actions must satisfy (4) the *plurality conditions*. But if actions satisfy the plurality conditions and the agents could have done otherwise voluntarily, intentionally, and rationally, then the agents could have done otherwise; and so they had (5) *alternative possibilities*. Therein lies the connection between UR and AP. If free will requires ultimate responsibility in the sense of UR, then at least *some* actions in our life histories (“will-setting actions”) must be such that we could have done otherwise with respect to them. Note, however, that this argument from (1) free will to (5) alternative possibilities (AP) is not direct. It goes *through* (2) ultimate responsibility (UR), (3) will-setting and (4) plurality; and UR is the key to it, since it is UR that implies will-setting and plurality. If we are to be ultimately responsible for our own *wills*, some of our actions must be such that we could have done otherwise, *because* some of them must have been such that we could have done otherwise voluntarily, intentionally, and rationally.

UR thus entails both indeterminism *and* alternative possibilities or AP. But it entails them by different argumentative routes. Two separate regresses are involved. (I call this the “dual regress of free will.”) The first regress begins with the requirement (of UR) that agents must be responsible by virtue of past voluntary actions for anything that is a *sufficient cause* of their actions. Stopping this regress requires that if agents are to have free will, some actions in their life histories must be *undetermined* (must lack sufficient causes). The second regress begins with the requirement that agents be responsible by virtue of past voluntary actions for anything that is a sufficient *motive* for their actions. Stopping this regress requires that some actions in an agent’s life history must be will-setting (so they do not have sufficient motives already set) and hence must satisfy the plurality conditions and hence AP.

The first of these two regresses results from the requirement that we be ultimate sources of our *actions*, the second from the requirement that we be ultimate sources of our *wills* (to perform those actions). If the second requirement were not added, we might have a world in which all the will-setting was done by someone or something other than the agents themselves, as in the imagined world in which all the will-setting was done by God. Agents in such a world might be unhindered in the pursuit of their purposes or ends, but it would never be “up to them” what *purposes* or ends they pursued. To have free will therefore is to be the ultimate designer of one’s own purposes or ends or goals. And if we are to be the ultimate designers of our own purposes or ends, there must be *some* actions in our life histories that are will-setting, plural voluntary *and* undetermined by someone or something else.

These undetermined, will-setting actions are the “*self-forming actions*,” or SFAs required by UR mentioned earlier. They would be the actions in our lives by which we ultimately *form* our character and motives and make ourselves into the kinds of persons we are.

## 6 The Intelligibility Problem: Is Libertarian Free Will Possible?

Can we make sense of a free will that requires Ultimate Responsibility of the kind described in the previous section? Can we really be the ultimate designers of our own ends and purposes? There are many skeptics about free will who think not. They argue that being the *ultimate* source of one’s will and actions is an incoherent and impossible ideal, since it would require us to be “prime movers unmoved” or “uncaused causes of ourselves” – “the best self-contradiction that has been conceived so far,” as Friedrich Nietzsche put it. Ultimate Responsibility or UR requires that there be some acts in our life

histories that do not have sufficient causes or motives. But how could acts having neither sufficient causes nor motives be free and responsible actions?

This question brings us to the second part of the modern attack on libertarian free will. It is one thing to offer arguments showing that free will is not compatible with determinism (and hence to address the “Compatibility Problem”). It is quite another thing to answer charges that an incompatibilist free will requiring ultimate responsibility is *intelligible* or *possible* and can be reconciled with modern scientific views of human beings. This is the “Intelligibility Problem” about libertarian free will; and it is in many ways even more difficult than the Compatibility Problem.

The culprit in the case of the Intelligibility Problem is not determinism, but *indeterminism*. For the Intelligibility Problem is related to an ancient dilemma noted earlier: if free will is not compatible with determinism, it does not seem to be compatible with indeterminism either. The arguments to show this have been made since ancient times. An undetermined or chance event, it is said, occurs spontaneously and is not controlled by anything, hence not controlled by the agent. To cite an example mentioned earlier, if a choice occurred by virtue of a quantum jump or other undetermined event in one’s brain it would seem a fluke or accident rather than a responsible choice. Such undetermined events occurring in our brains or bodies would not seem to enhance our freedom and control over our actions, but rather diminish our freedom and control.

Or we could put the Intelligibility Problem in another way that goes a little deeper. If my free choice is really undetermined, that means I could have made a different choice *given exactly the same past* right up to the moment when I did choose. That is what indeterminism and probability mean: given exactly the same past, different outcomes (“forking paths”) are possible. Imagine, for example, that John had been deliberating about where to spend his vacation, in Hawaii or Colorado, and after much thought and deliberation, had decided he preferred Hawaii and chose it. If the choice was undetermined, then exactly the same deliberation, the same thought processes, the same beliefs, desires, and other motives – not a sliver of difference – that led up to John’s favoring and choosing Hawaii over Colorado, might by chance have issued in his choosing Colorado instead. That is very strange. If such a thing happened it would seem a fluke or accident, like that quantum jump in the brain just mentioned, not a rational choice. Since John had come to favor Hawaii and was about to choose it, when by chance he chose Colorado, he might well wonder what went wrong and perhaps consult a neurologist.

One may at first think that there must be some way around the conclusion that if a choice is undetermined (like John’s choice just described), then the agent must have been able to choose otherwise “*given exactly the same past.*” But in fact there is no easy way around this conclusion. For indeterminism,

which is the denial of determinism, *does* mean “different possible futures, given the same past.” In the diagram of forking paths in section 1, the single line going back into the past is just that: a single line indicating “same past”; while the multiple lines going into the future represent “different possible futures.” By contrast, determinism means only one line into the future. If John is really free to choose different options at any time during his deliberation, and his choice is not determined, then he must be able to choose *either* path (Hawaii or Colorado), given the *same* past up to the moment when he chooses.

You can’t cheat here and say “If the past had been just a *tiny bit* different, then John might have sensibly and rationally chosen differently (chosen Colorado instead).” *Determinists* and *compatibilists* can say this. For they insist that John might have *sensibly* and *rationally* chosen otherwise only if the past had been different in some way (however small the difference). For example, if John had had a few different desires and beliefs or had reasoned a little differently, he might have come to favor Colorado and chosen it instead of Hawaii. But persons who believe free choices cannot be determined (as libertarians do) must say John may have chosen different possible futures, given the same entire past, including his psychological and physical history up to the moment he did choose. And this does seem to make his choosing otherwise (choosing Colorado) arbitrary and irrational in the same circumstances in which he actually came to favor Hawaii and chose it. You can see why many people have argued that undetermined free choices, of the kind libertarians demand, would be “arbitrary,” “capricious,” “random,” “irrational,” “uncontrolled,” and “inexplicable,” and not really free and responsible choices at all.

Defenders of libertarian free will, according to their critics, have a dismal record of answering such charges. Realizing that free will cannot merely be indeterminism or chance, libertarians have appealed to various unusual forms of agency or causation to make up the difference. For example, Immanuel Kant said we cannot explain free will in scientific and psychological terms, even though we require it for belief in morality. To account for free will, we have to appeal to the agency of what Kant called a “noumenal self” outside space and time that could not be studied in scientific terms. Many other respectable philosophers continue to believe that only some sort of appeal to mind/body dualism, of the kind associated with Descartes, can make sense of free will. Science might tell us there was indeterminacy or a place for causal gaps in the brain, but a non-material self or soul, or what Nobel physiologist John Eccles calls a “transempirical power center,” would have to fill the causal gaps left by physical causes by intervening in the natural order. The most popular appeal among libertarians today is to a special kind of *agent- or immanent causation* that cannot be explained in terms of the ordinary modes

of causation in terms of events familiar to the sciences. Free and responsible actions are not determined by prior events, according to this “agent-causation” view, but neither do free actions occur merely by chance. They are caused by the *agent* (a substance) in a way that transcends and cannot be explained in terms of ordinary modes of causation by events or states of affairs involving the agent.

I call these familiar libertarian strategies for making sense of free will “extra factor” strategies. The general idea behind all such strategies is not hard to understand: Since indeterminism means that an agent might act one way or in a different way, given exactly the same past, which would seem to include all the *same prior mental and physical events*, some “extra” kind of causation or agency must be postulated over and above the natural flow of events to account for the agent’s going one way rather than another. In short, some additional factor must be involved to “tip the balance.” It is this line of thought that has led libertarians through the centuries to postulate extra factors, such as immaterial causes, noumenal selves, transempirical power centers, non-event agent causes, prime movers unmoved, and so on, to explain free choices. And these postulates have in turn brought down on libertarians charges of obscurantism or mystery or “panicky metaphysics” from their critics.

Now it may be that some extra factors of the kinds just mentioned (or some others) *are* necessary to make sense of libertarian free will. Most libertarians today believe, for example, that some notion of “agent-causation” or causation by a substance that does not consist in causation in terms of events or states of affairs involving the agent, is required to make sense of free will. And this agent-causation view is ably defended by a number of recent philosophers, including Roderick Chisholm, Timothy O’Connor, Randolph Clarke, William Rowe, and others. But I happen to agree with other libertarians about free will, such as Peter van Inwagen and Carl Ginet, that “extra factor” strategies – including agent-causation theories – do not solve the problems about indeterminism they are suppose to solve and create further mysteries of their own. Moreover, “extra factor” strategies have tended to reinforce the widespread criticism that libertarian notions of free will requiring indeterminism are mysterious and have no place in the modern scientific picture of the world.

So my own belief is that, if we are going to make progress on the Intelligibility Problem about libertarian free will, we must strike out in new directions, trying to avoid to the degree possible appeals to extra factor strategies, including special forms of agent-causation, and appealing to such extra factors *only* if we cannot possibly avoid them. But doing this, I believe, means rethinking issues about indeterminism and responsibility, and hence libertarian free will, from the ground up – a task to which I now turn.

## 7 Indeterminism and Responsibility

The first step in this rethinking about the Intelligibility Problem is to note that indeterminism does not have to be involved in *all* acts done “of our own free wills” for which we are ultimately responsible, as noted earlier. All free acts do not have to be undetermined on the libertarian view, but only those acts by which we made ourselves into the kinds of persons we are, namely the “will-setting” or “self-forming actions” (SFAs) that are required for ultimate responsibility.

Now I believe these undetermined self-forming actions or SFAs occur at those difficult times of life when we are torn between competing visions of what we should do or become. Perhaps we are torn between doing the moral thing or acting from ambition, or between powerful present desires and long-term goals, or we are faced with difficult tasks for which we have aversions. In all such cases, we are faced with competing motivations and have to make an effort to overcome temptation to do something else we also strongly want. There is tension and uncertainty in our minds about what to do at such times, I suggest, that is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium – in short, a kind of “stirring up of chaos” in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. The uncertainty and inner tension we feel at such soul-searching moments of self-formation is thus reflected in the indeterminacy of our neural processes themselves. What we experience internally as uncertainty about what to do on such occasions would then correspond physically to the opening of a window of opportunity that temporarily screens off complete determination by influences of the past.

When we do decide under such conditions of uncertainty, the outcome would not be determined because of the preceding indeterminacy – and yet the outcome can be willed (and hence rational and voluntary) either way owing to the fact that in such self-formation, the agents’ prior wills are divided by conflicting motives. Consider a businesswoman who faces such a conflict. She is on her way to an important meeting when she observes an assault taking place in an alley. An inner struggle ensues between her conscience, to stop and call for help, and her career ambitions, which tell her she cannot miss this meeting. She has to make an effort of will to overcome the temptation to go on. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not *allow* her effort to succeed. And this is due to the fact that, while she willed to overcome temptation, she also willed to fail, for quite different and incommensurable reasons. When we, like the woman, decide in such circumstances, and the indeterminate efforts we are making become determinate choices, we *make* one set of

competing reasons or motives prevail over the others then and there *by deciding*.

Now let us add a further piece to the puzzle. Just as indeterminism need not undermine rationality and voluntariness of choices, so indeterminism in and of itself need not undermine control and responsibility. Suppose you are trying to think through a difficult problem, say a mathematical problem, and there is some indeterminacy in your neural processes complicating the task – a kind of chaotic background. It would be like trying to concentrate and solve a problem, say a mathematical problem, with background noise or distraction. Whether you are going to succeed in solving the problem is uncertain and undetermined because of the distracting neural noise. Yet, if you concentrate and solve the problem nonetheless, we have reason to say you did it and are responsible for it, even though it was undetermined whether you would succeed. The indeterministic noise would have been an obstacle that you overcame by your effort.

There are numerous examples supporting this point, where indeterminism functions as an obstacle to success without precluding responsibility. Included among these examples are the Austin-style examples discussed in section 5. Recall the assassin who is trying to shoot the prime minister, but might miss because of some undetermined events in his nervous system that may lead to a jerking or wavering of his arm. If the assassin does succeed in hitting his target, despite the indeterminism, can he be held responsible? The answer is clearly yes because he intentionally and voluntarily succeeded in doing what he was *trying* to do – kill the prime minister. Yet his action, killing the prime minister, was undetermined. Indeterminism, it would appear, does not necessarily rule out responsibility.

Here is another example: A husband, while arguing with his wife, in a fit of rage swings his arm down on her favorite glass-top table top intending to break it. Again, we suppose that some indeterminism in his outgoing neural pathways makes the momentum of his arm indeterminate, so that it is undetermined whether the table will actually break right up to the moment when it is struck. Whether the husband breaks the table is undetermined and yet he is clearly responsible, if he does break it. (It would be a poor excuse to offer his wife, if he claimed: “Chance did it, not me.” Though indeterminism was involved, chance didn’t do it, he did.) In this example, as in the previous one, the agent can be held responsible for an action even though the action was undetermined.

Now these examples – of the mathematical problem, the assassin and the husband – are not all we want for free will, since they do not amount to genuine exercises of self-forming actions (SFAs), like the businesswoman’s, where the will is divided between conflicting motives. The businesswoman wants to help the victim, but she also wants to go on to her meeting. By

contrast, the assassin's will is not equally divided. He wants to kill the prime minister, but he does not also want to fail. (If he fails therefore, it will be *merely* by chance.) Yet these examples of the assassin, the husband and the like, while they do not tell us all we want to know about free will, do provide some clues about what free will requires. To go further, we have to add some additional twists.

## 8 Parallel Processing

Imagine in cases of conflict characteristic of self-forming actions or SFAs, like the businesswoman's, that the indeterministic noise which is providing an obstacle to her overcoming temptation is not coming from an external source, but has its source in her own will, since she also deeply desires to do the opposite. To understand how this could be, imagine that two crossing recurrent neural networks are involved in the brain, each influencing the other, and representing her conflicting motivations. (Recurrent neural networks are complex networks of interconnected neurons in the brain circulating impulses in feedback loops that are generally involved in higher-level cognitive processing.) The input of one of these neural networks consists in the woman's reasons for acting morally and stopping to help the victim; the input of the other network comprises her ambitious motives for going on to her meeting.

The two networks are connected so that the indeterminism that is an obstacle to her making one of the choices is present because of her simultaneous conflicting desire to make the other choice – the indeterminism thus arising from a tension-creating conflict in the will, as we said. This conflict, as noted earlier, would be reflected in appropriate regions of the brain by movement away from thermodynamic equilibrium. The result would be a stirring up of chaos in the neural networks involved. Chaos in physical systems is a phenomenon in which very small changes in initial conditions are magnified so that they lead to large and unpredictable changes in the subsequent behavior of a system. You may have heard the popular illustration of chaos in which the fluttering of a butterfly's wings in South America initiates a chain of events that affects the weather patterns of North America. Such popular examples may be an exaggeration. But chaotic phenomena, in which small changes lead to large effects, are now known to be far more common in nature than previously believed; and they are particularly common in living things. There is growing evidence that chaos plays a role in the information processing of the brain, providing some of the flexibility that the nervous system needs to adapt creatively – rather than in predictable or rigid ways – to an ever-changing environment.



Now determinists are quick to point out that chaos, or chaotic behavior, in physical systems, though *unpredictable*, is usually deterministic and does not itself imply genuine indeterminism in nature. But some scientists have suggested that a combination of chaos and quantum physics might provide the genuine indeterminism one needs. If the processing of the brain does “make chaos in order to make sense of the world” (as one recent research paper puts it), then the resulting chaos might magnify quantum indeterminacies in the firings of individual neurons so that they would have large-scale indeterministic effects on the activity of neural networks in the brain as a whole. If chaotic behavior were thus enhanced in these neural networks by tension-creating conflict in the will, the result would be some significant indeterminism in the cognitive processing of each of the competing neural networks.

In such circumstances, when either of the competing networks “wins” (or reaches an activation threshold, which amounts to choice), it would be like your solving the mathematical problem by overcoming the background indeterministic noise created by the presence of the competing network. And just as when you solved the mathematical problem by overcoming the distracting noise, one can say you did it and are responsible for it, so one can also say this, I would argue, in the present case, *whichever outcome is chosen*. For the neural pathway through which the woman does succeed in reaching a choice threshold will have overcome the obstacle in the form of indeterministic noise generated by the presence of the other competing network.

Note that, under such conditions, the choice the woman might make either way will not be “inadvertent,” “accidental,” “capricious,” or “merely random” (as critics of indeterminism say) because the choice will be *willed* by the woman either way when it is made, and it will be done for *reasons* either way – reasons that she then and there *endorses*. For, let us recall that in SFAs, the agent’s will is divided and the agent has strong reasons or motives for making *either* choice. So when she decides, she endorses one set of competing reasons over the other as the one she will act on. But *willing* what you do in this way, and doing it for *reasons* that you endorse, are conditions usually required to say something is done “on purpose,” rather than accidentally, capriciously, or merely by chance. Moreover, these conditions taken together (that the choices were willed either way, were done for reasons and the agents endorsed them) rule out each of the reasons we have for saying that agents act, but do not have *control* over their actions. The businesswoman’s choice either way, for example, will not have been made accidentally or inadvertently or by mistake, nor need it have been the result of coercion (no one was holding a gun to her head, for example) or the result of control by other agents. Of course, for undetermined SFAs, agents do not control or determine which choice outcome will occur *before* it occurs. But it does not follow, because one does not control

or determine which of a set of outcomes is going to occur before it occurs, that one does not control or determine which of them occurs, *when* it occurs. When the above conditions for SFAs are satisfied, agents exercise control over their future lives *then and there* by deciding.

As a consequence, they have what I call *plural voluntary control* over their options in the following sense: Agents have plural voluntary control over a set of options (such as the woman's choosing to help the victim or to go on to her meeting), when they are able to bring about *whichever* of the options they will, *when* they will to do so, *for* the reasons they will to do so, *on* purpose, rather than accidentally or by mistake, *without* being coerced or compelled in doing so or willing to do so, or otherwise controlled in doing or willing to do so by any other agents or mechanisms. Each of these conditions can be satisfied for SFAs, like the businesswoman's, as I have described them. The conditions can be summed up by saying that the agents can choose either way *at will*. In other words, the choices are "will-setting": We set our wills one way or the other in the *act* of deciding itself, and not before.

Note also that this account of self-forming choices or SFAs amounts to a kind of "doubling" of the mathematical problem. It is as if an agent faced with such a self-forming choice is *trying* or making an effort to solve *two* cognitive problems at once, or to complete two competing (deliberative) tasks at once – in our example, to make a moral choice and to make a conflicting self-interested choice (corresponding to the two competing neural networks involved). Each task is being thwarted by the indeterminism generated by the presence of the competing network, so it might fail. But if it succeeds, then the agents can be held responsible because, as in the case of solving the mathematical problem, the agents will have succeeded in doing what they were knowingly and willingly trying to do.

Recall the assassin and the husband. Owing to indeterminacies in their neural pathways, the assassin might miss his target or the husband fail to break the table. But if they *succeed*, despite the probability of failure, they are responsible, because they will have succeeded in doing what they were trying to do. And so it is, I suggest, with self-forming choices (SFAs) like the businesswoman's, except that in the case of self-forming choices, *whichever way the agents choose* they will have succeeded in doing what they were trying to do because they were simultaneously trying to make both choices, and one is going to succeed. Their failure to do one thing is not a *mere* failure, but a voluntary succeeding in doing the other.

Does it make sense to talk about the agent's trying to do two competing things at once in this way, or to solve two cognitive problems at once? Well, we now know that the brain is a "parallel processor"; it can simultaneously process different kinds of information relevant to tasks such as perception or recognition through different neural pathways. Such a capacity, I believe, is

essential to the exercise of free will. In cases of self-formation (SFAs), agents are simultaneously trying to resolve plural and competing cognitive tasks. They are, as we say, of two minds. Yet they are not two separate persons. They are not dissociated from either task. The businesswoman who wants to go back to help the victim is the same ambitious woman who wants to go to her meeting and make a sale. She is torn inside by different visions of who she is and what she wants to be, as we all are from time to time. But this is the kind of complexity needed for genuine self-formation and free will. And when she succeeds in doing one of the things she is trying to do, she will endorse that outcome as *her* resolution of the conflict in her will, voluntarily and intentionally, not by accident or mistake.

## 9 Responsibility, Luck, and Chance

You may find all this interesting and yet still find it hard to shake the intuition that if choices are undetermined, they *must* happen merely by chance – and so must be “random,” “capricious,” “uncontrolled,” “irrational,” and all the other things usually charged. Such intuitions are deeply ingrained. But if we are going to understand free will, I think we must break old habits of thought supporting such intuitions and learn to think in new ways.

The first step is to question the intuitive connection in people’s minds between “indeterminism’s being involved in something” and “its happening merely as a matter of chance or luck.” “Chance” and “luck” are terms of ordinary language that carry the meaning of “its being out of my control.” So using them already begs certain questions. Whereas “indeterminism” is a technical term that merely rules out *deterministic* causation, though not causation altogether. Indeterminism is consistent with nondeterministic or probabilistic causation, where the outcome is not inevitable. It is therefore a mistake (in fact, one of the most common in debates about free will) to assume that “undetermined” means “uncaused” or “*merely* a matter of chance.”

Here is another source of misunderstanding. Since the outcome of the businesswoman’s effort (the choice) is undetermined up to the last minute, we may have the image of her first making an effort to overcome the temptation to go on to her meeting and then at the last instant “chance takes over” and decides the issue for her. But this is a mistaken image. On the view just presented, one cannot separate the indeterminism and the effort of will, so that *first* the effort occurs *followed* by chance or luck (or vice versa). One must think of the effort and the indeterminism as fused; the effort *is* indeterminate and the indeterminism is a property of the effort, not something separate that occurs after or before the effort. The fact that the effort has this property of being indeterminate does not make it any less the woman’s *effort*. The complex

recurrent neural network that realizes the effort in the brain is circulating impulses in feedback loops and there is some indeterminacy in these circulating impulses. But the whole process is her effort of will and it persists right up to the moment when the choice is made. There is no point at which the effort stops and chance “takes over.” She chooses as a result of the effort, even though she might have failed. Similarly, the husband breaks the table as a result of his effort, even though he might have failed because of the indeterminacy. (That is why his excuse, “chance broke the table, not me,” is so lame.)

Just as expressions like “she chose *by chance*” can mislead us in such contexts, so can expressions like “she got lucky.” Recall that one might say of the assassin and husband “they got lucky” in killing the prime minister and breaking the table because their actions were undetermined. Yet the surprising thing is that we can still say the assassin and husband were *responsible* if they succeeded in killing the prime minister and breaking the table. So we should ask ourselves the following question: why is it wrong to say “he got lucky, *so he was not responsible*” in the cases of the husband and the assassin? For it *is* wrong to say this since they did get lucky and yet they were *still* responsible. (Imagine the assassin’s lawyer arguing in the courtroom that his client is not guilty because his killing the prime minister was undetermined and might therefore have failed by chance. Would such a defense succeed?)

The first part of an answer as to why the assassin and husband are still responsible has to do with the point made earlier about “luck” and “chance.” These two words have question-begging implications in ordinary language that are not necessarily implications of “indeterminism” (which implies only the absence of deterministic causation). The core meaning of “he got lucky” in the assassin and husband cases, which *is* implied by indeterminism, is that “he succeeded *despite the probability or chance of failure*”; and this core meaning does not imply lack of responsibility, *if he succeeds*. If “he got lucky” had other meanings in these cases that are often associated with “luck” and “chance” in ordinary usage, the inference “he got lucky so he was not responsible” would not fail for the husband and assassin, as it clearly does. For example, if “luck” in these cases meant the outcome was not his doing, or occurred by mere chance, or he was not responsible, then the inference “he got lucky so he was not responsible” would hold for the husband and assassin. But the point is that these further meanings of “luck” and “chance” do not follow *from the mere presence of indeterminism*.

The second reason why the inference “he got lucky, so he was not responsible” does not work in the cases of the assassin and the husband is that *what* they succeeded in doing was what they were *trying* and wanting to do all along (kill the minister and break the table respectively). The third reason is that *when* they succeeded, their reaction was not “Oh dear, that was a mistake,

an accident – something that *happened* to me, not something I *did*.” Rather they *endorsed* the outcomes as something they were trying and wanting to do all along, knowingly and purposefully, not by mistake or accident.

But these conditions are satisfied in the businesswoman’s case as well *either way* she chooses. If she succeeds in choosing to return to help the victim (or in choosing to go on to her meeting) (i) she will have “succeeded *despite the probability or chance of failure*,” (ii) she will have succeeded in doing what she was *trying* and *wanting* to do all along (she wanted both outcomes very much, but for different reasons, and was trying to make those reasons prevail in both cases), and (iii) when she succeeded (in choosing to return to help) her reaction was not “Oh dear, that was a mistake, an accident – something that happened to me, not something I did.” Rather she *endorsed* the outcome as something she was trying and wanting to do all along; she recognized the choice as her resolution of the conflict in her will. And if she had chosen to go on to her meeting she would have endorsed that outcome, recognizing it as her resolution of the conflict in her will.

## 10 Choice, Agency, Efforts, and Causes: Further Objections Considered

Perhaps we are begging the question by assuming the outcomes of the woman’s efforts are *choices* to begin with. If indeterminism is involved in a process (such as the woman’s deliberation) so that its outcome is undetermined, one might argue that the outcome must merely *happen* and therefore cannot be somebody’s *choice*. But there is no reason to assume such a claim is true. A choice is the formation of an intention or purpose to do something. It resolves uncertainty and indecision in the mind about what to do. Nothing in such a description implies that there could not be some indeterminism in the deliberation and neural processes of an agent preceding choice corresponding to the agent’s prior uncertainty about what to do. Recall from the preceding arguments that the presence of indeterminism does not mean the outcome happened *merely* by chance and *not* by the agent’s effort. Self-forming choices are undetermined, but not uncaused. They are caused by the agent’s efforts.

Well, perhaps indeterminism does not undermine the idea that something is a *choice* simply, but rather that it is the *agent’s* choice. This objection raises important questions about agency. What makes the woman’s choice her own on the above account is that it results from *her* efforts and deliberation, which in turn are causally influenced by her reasons and her intentions (for example, her intention to resolve indecision in one way or another). And what makes these efforts, deliberation, reasons, and intentions *hers* is that they are embedded in a larger motivational system realized in her brain in terms of which

she defines herself as a practical reasoner and actor. A choice is the agent's when it is produced intentionally by efforts, by deliberation and by reasons that are part of this self-defining motivational system and when, in addition, the agent *endorses* the new intention or purpose created by the choice into that motivational system as a further purpose to guide *future* practical reasoning and action.

Another concern that has been raised about the above account of libertarian free will is that we are not introspectively aware of making dual efforts and performing multiple cognitive tasks in such choice situations. But I am not claiming that agents are conscious of making dual efforts. What they are introspectively conscious of is that they are trying to decide about which of two options to choose and that either choice is a difficult one because there are resistant motives pulling them in different directions that will have to be overcome, whichever choice is made. In such introspective conditions, I am theorizing that what is actually going on underneath is a kind of parallel processing in the brain that involves separate efforts or endeavors to resolve competing cognitive tasks. The point is that introspective evidence does not give us the whole story about free will. If we stay on the surface and just consider what our immediate experience tells us, free will, I believe, is bound to appear mysterious, as it has appeared to so many people through the centuries. To unravel its mysteries, we have to consider what might be going on behind the scenes.

It is now widely believed, for example, that parallel processing takes place in the brain in such cognitive phenomena as visual perception. The theory is that the brain separately processes different features of the visual scene, such as object and background, through distinguishable and parallel, though interacting, neural pathways or streams. Suppose someone objected that we are not introspectively aware of such distributed processing in ordinary cases of perception. That would hardly be a decisive objection to this new theory of vision. For the claim is that this is what we are doing in visual perception, not necessarily that we are introspectively aware of doing it. And I am making a similar claim about free will. What is needed is a *theory* about what might be going on when we exercise free will, not merely a description of what we immediately experience.

It has also been objected that it is irrational to make efforts to do incompatible things. I concede that in most ordinary situations it is. But building on suggestions made by theorists of action, such as Michael Bratman, I argue that there are special circumstances in which it is not irrational to make competing efforts: These include circumstances in which: (i) we are deliberating between competing options; (ii) we intend to choose one or the other, but cannot choose both; (iii) we have powerful motives for wanting to choose each of the options for different and incommensurable reasons; (iv) there is

a consequent resistance in our will to either choice, so that (v) if either choice is to have a chance of being made, effort will have to be made to overcome the temptation to make the other choice; and, most importantly, (vi) we want to give each choice a fighting chance of being made because the motives for each choice are important to us. The motives for each choice define in part what sort of person we are; and we would taking them lightly if we did not make an effort in their behalf. These conditions are, of course, the conditions of SFAs.

But perhaps the deepest concern about the above theory remains the concern about *chance*. If chance is involved in decision making, we somehow think of chance as deciding the issue, like spinning a wheel to select an outcome. As noted earlier, that worry sends us scurrying around looking for extra factors, other than prior events or happenings, to tip the balance to one choice or the other, such as an immaterial agent or non-event agent cause. But there is an alternative way to think about the way that indeterminism might be involved in free choice that first occurred to me twenty-five years ago, a way that avoids these familiar libertarian stratagems and requires a transformation of perspective.

Think, instead, of the indeterminism involved in free choice as an *ingredient* in a larger goal-directed or teleological process or activity, in which the indeterminism functions as a *hindrance* or *obstacle* to the attainment of the goal. If you reflect for a moment, you will see that this is what the account of free will presented earlier is actually doing. Here is an example from another modern scientific theory of relevance to free will – namely, information theory. Consider the sending of a message in Morse code. The sender taps out the message in dots and dashes, representing letters. The pulses travel electrically over lines to the receiver where they are reproduced. Now, there may be interference due to noise or static in the electrical lines so that the message does not get through, or a distorted message gets through. In that case we have what information theorists call “equivocation” rather than mere noise. The message is too garbled to read. If the message does get through, however, despite the electrical noise or static, then the goal of the message sender is realized. Now if the noise in the electrical lines were the result of indeterminism or chance, whether the message gets through would be undetermined. Yet if the undetermined electrical noise or static was not great enough to cause equivocation, the goal of the process would be realized, despite the interference (the message would get through despite the indeterminism).

In a similar fashion, the idea is not to think of the indeterminism involved in free choices as a cause *acting on its own*, but as an ingredient in a larger goal-directed or teleological process or activity in which the indeterminism functions as a hindrance or obstacle to the attainment of the

goal. This is the role suggested for indeterminism in the efforts preceding undetermined SFAs. These efforts are temporally extended goal-directed activities in which indeterminism is a hindering or interfering element, like the noise or static in the message transmission example. The choices, or SFAs, that result from these temporally extended activities or efforts thus do not pop up out of nowhere, even though they are undetermined. They are the *achievements* of goal-directed activities of the agent that might have failed, but did not.

Note that, if indeterminism or chance does play this kind of interfering role in a larger process leading to choice, the indeterminism or chance need not be the *cause* of the choice that is actually made. This follows from a general point about probabilistic causation. A vaccination may hinder or lower the probability that I will get a certain disease, so it is causally relevant to the outcome. But if I get the disease despite it, the vaccination is not the *cause* of my getting the disease, though it was causally relevant, because its role was to *hinder* that effect. The causes of my getting the disease, by contrast, are those causally relevant factors (such as the infecting virus) that significantly *raised* the probability of its occurrence. Similarly, in the case of the businesswoman's choice, the causes of the choice she does make (the moral choice or the ambitious choice) are those causally relevant factors that significantly *raised* the probability of making *that* choice from what it would have been if those factors had not been present, such as her reasons and motives for making that choice rather than the other, her conscious awareness of these reasons and her deliberative efforts to overcome the temptations to make the contrary choice. The presence of the indeterminism lowers the probability that the choice will result from these reasons, motives, and efforts from what that probability would have been if there had been no competing motives or efforts and hence no interfering indeterminism.

Since those causally relevant features of the agent, which *can* be counted among the causes of the woman's choice, are *her* reasons or motives, *her* conscious awareness and *her* deliberative efforts, we can also say that she is the cause of the choice by virtue of making the efforts for the reasons and succeeding. The indeterminism or chance (like the vaccination) was causally relevant to the outcome, but it was not the cause. This explains why the husband's excuse was so lame when he said "Chance broke the table, not me." While chance was causally involved, chance was not the cause of the table's breaking. The cause was his effort to break the table by swinging his arm down on it. The chance merely made it uncertain whether that larger goal-directed activity would succeed. And so it is, I suggest, with the efforts leading to self-forming choices. These efforts, of course, are mental activities realized in the higher cognitive processing of the brain rather than in overt actions such as the swinging of an arm. But the SFAs that result from



these mental efforts are nonetheless also the achievements of goal-directed activities that might have failed due to chance, but did not, just as the husband's effort to break the table by swinging his arm might have failed due to chance, but did not.

But can't we say that it is a "matter of chance" whether one of these efforts leading to SFAs succeeds or not? For isn't it true that whether or not an effort succeeds in producing a choice depends on whether certain undetermined neurons involved in the agent's cognitive processing fire or do not fire (perhaps within a given time frame)? And whether these neurons fire or not is by hypothesis undetermined, is it not, and therefore not under the control of the agent? Well, yes, we *can* say all of these things: whether an effort succeeds *does* depend upon whether certain undetermined neurons fire or not; and whether these neurons fire is not under the control of the agent; and we can consequently say it is a matter of chance whether the efforts leading to SFAs succeed or not.

But the really astonishing thing is that, while all these things can be truly said, it *does not follow* that the agent is not *responsible* for the choice, *if* the effort succeeds. For, consider the husband swinging his arm down on the table. It is *also* true in his case that whether or not his effort to break the table succeeds "depends" on whether certain neurons in his arm fire or do not fire; and it is *also* true in his case that whether these neurons fire or not is undetermined and therefore not under his control; and we can *also* consequently say in the husband's case that it is a "matter of chance" whether or not he succeeds in breaking the table. Yet, even though we can say all this, it does not follow that he is not responsible for breaking the table, *if* his effort succeeds. Astonishing indeed! But this is the kind of surprising result one gets when indeterminism or chance plays an interfering or hindering role in larger goal-directed activities, such as efforts to do certain things that may succeed or fail.

It is well to meditate on this: We tend to reason that if an outcome (breaking a table *or* making a choice) depends on whether certain neurons fire or not (in the arm *or* in the brain), then the agent must be able to *make* those neurons fire or not, if the agent is to be responsible for the outcome. In other words, we think we have to crawl down to the place where the indeterminism originates (in the individual neurons) and *make* them go one way or the other. We think we have to become originators at the micro-level and tip the balance that chance leaves untopped, if we (and not chance) are to be responsible for the outcome. And we realize, of course, that we can't do that. But we don't have to. It's the wrong place to look. We don't have to micro-manage our individual neurons one by one to perform purposive actions. In fact, we do not have such micro-control over our neurons *even when we perform ordinary actions* such as swinging an arm down on a table.

What we need when we perform purposive activities, mental or physical, is rather macro-control of processes involving many neurons – complex processes that may succeed in achieving their goals despite the interfering effects of some recalcitrant neurons. We don't micro-manage our actions by controlling each individual neuron or muscle that might be involved. We don't know enough about neurology or physiology to do that; and it would be counter-productive to try. But that does not prevent us from macro-managing our purposive activities (whether they be mental activities such as practical reasoning, or physical activities, such as arm-swingings) and being responsible when those purposive activities attain their goals.

## 11 Responsibility and Control: Three Assassins

But does not the presence of indeterminism or chance at least *diminish* the control persons have over their choices or actions? And would that not affect their responsibility? (This is another way in which objections about chance and luck have often been raised against libertarian views of free will.) Is it not the case that the assassin's control over whether the prime minister is killed (his ability to realize his purposes or what he is trying to do) is lessened by the undetermined impulses in his arm – and so also for the husband and his breaking the table? The answer is yes, again. But the further surprising point worth noting – a point that I think is so often missed – is that *diminished control* in such circumstances *does not entail diminished responsibility* when the agents succeed in doing what they are trying to do. Ask yourself this question: Is the assassin less guilty of killing the prime minister, if he did not have complete control over whether he would succeed because of the indeterminism in his neural processes?

Suppose there were three assassins, each of whom killed a prime minister. Suppose one of them had a 50 percent chance of succeeding because of the indeterministic wavering of his arm. Another had an 80 percent chance, and the third a 100 percent chance. (With this third assassin there was no wavering at all; he was a young stud assassin.) Is one of these assassins less guilty than the other, *if they all succeed*? Should we say that one assassin deserves a hundred years in jail, the other eighty years and the third fifty years? Absurd. They are all equally guilty if they succeed. The diminished control in the assassins who had an 80 percent or a 50 percent chance does not translate into diminished responsibility when they succeed. Diminished control in such circumstances does not entail diminished responsibility. Imagine a lawyer for the 50 percent assassin arguing that his client was not guilty because the prime minister's dying as a result of what his client did was a "matter of chance." Therefore chance was the cause of the prime minister's death, not

his client. That would make the notorious “Twinkie Defense” look brilliant by comparison. (This was the defense offered by a lawyer in California that his client was not responsible because the client’s blood sugar was so high from having eaten too many Twinkies that he could not control his actions.)

There is an important further lesson here I believe about free will in general. We should concede that indeterminism, wherever it occurs, *does* diminish control over what we are trying to do and *is* a hindrance or obstacle to the realization of our purposes. But recall that in the case of the business-woman (and SFAs generally), the indeterminism that is admittedly diminishing her control over one thing she is trying to do (the moral act of helping the victim) *is coming from her own will* – from her desire and effort to do the opposite (go to her business meeting). And the indeterminism that is diminishing her control over the other thing she is trying to do (act selfishly and go to her meeting) is coming from her desire and effort to do the opposite (to be a moral person and act on moral reasons). In each case, the indeterminism *is* functioning as a hindrance or obstacle to her realizing one of her purposes – a hindrance or obstacle in the form of resistance within her will which has to be overcome by effort.

If there were no such hindrance – if there were no resistance in her will – she would indeed in a sense have “complete control” over one of her options. There would be no competing motives standing in the way of her choosing it and therefore no interfering indeterminism. But then also, she would not be free to rationally and voluntarily choose the *other* purpose because she would have no good competing reasons to do so. Thus, by *being* a hindrance to the realization of some of our purposes, indeterminism paradoxically opens up the genuine possibility of pursuing other purposes – of choosing or doing *otherwise* in accordance with, rather than against, our wills (voluntarily) and reasons (rationally). To be genuinely self-forming agents (creators of ourselves) – to have free will – there must at times in life be obstacles and hindrances in our wills of this sort that we must overcome.

I think libertarians about free will have traditionally tried to ignore this aspect of indeterminism. They knew indeterminism was required on their view, but assumed it could be entirely circumvented by special agencies. But hindrances and obstacles and resistance in the will are precisely what are needed for free will, which, like life itself, exists near the edge of chaos. If one were to put it in a religious perspective, this fact would be related to the problem of evil. There must be hindrances and obstacles to our choices and resistance in our own wills to be overcome, if we are to be capable of genuine self-formation and free will. Compare Evodius’s question to St Augustine (in Augustine’s classic work *On the Free Choice of the Will*) of why God gave us free will since it brings so much conflict, struggle and suffering into the world.

Yes, it does bring struggle, hindrances and resistance in our wills. But such things are necessary for genuine responsibility.

Of interest also is Kant's image, which I have used before, of the bird that is upset by the resistance of the air and the wind to its flight and so imagines that it could fly better if there were no air at all to resist it. But of course the bird would not fly better if there were no air. It would cease to fly at all. So it is with indeterminism in relation to free will. It provides resistance to our choices, but a resistance that is necessary if we are to be capable of true self-formation.

## **12 Conclusion: Complexity and "Being an Author of One's Own Story"**

In summary, I think the key to understanding the role of chance in free will is not to think of chance as a causal factor by itself, but rather to think of chance as an interfering ingredient in larger goal-directed processes. Viewing chance in this way is related to a peculiarly modern scientific way of understanding human agency that also has its roots in the ancient view of Aristotle. Agents, according to this modern conception with ancient roots, are to be conceived as *information-responsive complex dynamical systems*. Complex dynamical systems are the subject of "dynamical systems theory" and also of what is sometimes popularly called "complexity theory." They are systems (which are now known to be ubiquitous in nature) in which new *emergent* capacities arise as a result of greater complexity or as the result of movement away from thermodynamic equilibrium toward the edge of chaos. When these emergent capacities arise in complex dynamical systems, the systems as a whole impose novel constraints on the behavior of their parts that did not constrain the parts before the new complexity or disequilibrium was achieved. In such complex dynamical systems there is thus a reciprocal causal influence of wholes to parts and parts to wholes.

In the account of free will I have proposed, for example, it is a conflict in the larger motivational system of the agent taken as a whole – the self-network, as I have elsewhere called it – that stirs up chaos and amplifies indeterminism at the neuronal and synaptic levels. The larger whole or self-network thus stirs up chaos in its parts (neurons and networks of neurons), but the resulting amplified indeterminism in turn interferes with the goal-directed activities of the larger network. There is thus a mutual influence of wholes to parts and parts to wholes characteristic of complex dynamical systems. And emergent capacities are also involved. Only when creatures attain the kind of inner complexity capable of giving rise to conflicts in their wills, or motivational systems, between incommensurable values does

the capacity for self-formation characteristic of free will arise. So we are talking about a special kind of complex dynamical system that is information-responsive in highly complex ways, not seen in non-rational animals. The businesswoman, as I said, is torn inside by different visions of who she is and what she wants to be, as we all are from time to time. But this is just the kind of complexity needed for the novel capacity of genuine self-formation or free will to emerge.

Let me conclude with one final objection to the account of free will presented here, which is perhaps the most telling and has not yet been discussed. Even if one granted that persons, such as the businesswoman, could make genuine self-forming choices that were undetermined, isn't there something to the charge that such choices would be *arbitrary*? A residual arbitrariness seems to remain in all self-forming choices since the agents cannot in principle have sufficient or conclusive *prior* reasons for making one option and one set of reasons prevail over the other.

There is some truth to this objection also, but again I think it is a truth that tells us something important about free will. It tells us that every undetermined self-forming free choice is the initiation of what might be called a *value experiment* whose justification lies in the future and is not fully explained by past reasons. In making such a choice we say, in effect, "Let's try this. It is not required by my past, but it is consistent with my past and is one branching pathway in the garden of forking paths my life can now meaningfully take. Whether it is the right choice, only time will tell. Meanwhile, I am willing to take responsibility for it one way or the other."

It is worth noting that the term "arbitrary" comes from the Latin *arbitrium*, which means "judgment" – as in *liberum arbitrium voluntatis*, "free judgment of the will" (the medieval philosophers' designation for free will). Imagine a writer in the middle of a novel. The novel's heroine faces a crisis and the writer has not yet developed her character in sufficient detail to say exactly how she will act. The author makes a "judgment" about this that is not determined by the heroine's already formed past which does not give unique direction. In this sense, the judgment (*arbitrium*) of how she will react is "arbitrary," but not entirely so. It had input from the heroine's fictional past and in turn gave input to her projected future. In a similar way, agents who exercise free will are both authors of and characters in their own stories all at once. By virtue of "self-forming" judgments of the will (*arbitria voluntatis*) (SFAs), they are "arbiters" of their own lives, "making themselves" out of past that, if they are truly free, does not limit their future pathways to one.

Suppose we were to say to such persons: "But look, you didn't have sufficient or *conclusive* prior reasons for choosing as you did since you also had viable reasons for choosing the other way." They might reply. "True enough. But I did have *good* reasons for choosing as I did, which I'm willing to stand

by *and take responsibility for*. If these reasons were not sufficient or conclusive reasons, that's because, like the heroine of the novel, I was not a fully formed person before I chose (and still am not, for that matter). Like the author of the novel, I am in the process of writing an unfinished story and forming an unfinished character who, in my case, is myself."

### Further Reading

For more advanced discussion of the issues discussed in this chapter, see the collection of readings in Robert Kane (ed.), *The Oxford Handbook of Free Will* (Oxford: Oxford University Press, 2002). The following collections of essays also contain further readings on the issues about free will discussed in this chapter: Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press, 2003), Robert Kane (ed.), *Free Will* (Oxford: Blackwell Publishers, 2002), and Laura Ekstrom (ed.), *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press, 2001). Another collection of readings which deals specifically with libertarian accounts of free will is Timothy O'Connor (ed.), *Agents, Causes and Events: Essays on Free Will and Indeterminism* (Oxford: Oxford University Press, 1995).

The libertarian view of free will presented in this chapter is further developed in my book *The Significance of Free Will* (Oxford: Oxford University Press, 1996). Alternative accounts of libertarian free will may also be found in Timothy O'Connor, *Persons and Causes* (Oxford: Oxford University Press, 2000); Randolph Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford University Press, 2003); Carl Ginet, *On Action* (Cambridge: Cambridge University Press, 1990); Hugh McCann, *The Works of Agency* (Ithaca, NY: Cornell University Press, 1998), Stewart Goetz, "A Non-causal Theory of Agency," *Philosophy and Phenomenological Research* 49 (1988), 303–16; Laura Waddell Ekstrom, *Free Will* (Boulder, CO: Westview Press, 2000); David Hodgson "A Plain Person's Free Will," *Journal of Consciousness Studies* 12 (2005), 3–19; James Felt, *Making Sense of Our Freedom* (Ithaca, NY: Cornell University Press, 1994), and Thomas Pink, *Free Will: A Short Introduction* (Oxford: Oxford University Press, 2004). O'Connor and Clarke defend sophisticated modern versions of the "agent-causation theory" of libertarian free will, which was mentioned in the chapter. Ginet, McCann, and Goetz defend what are called "simple indeterminist" or "noncausalist" accounts of libertarian free will. The view defended in this chapter is often called a "causal indeterminist" or "event-causal" view of libertarian free will to distinguish it from "agent-causation" and "simple indeterminist" views. Ekstrom also defends a causal indeterminist view, though different than the one defended in this chapter. The libertarian views defended by Hodgson, Felt, and Pink are not easily fitted into any of these three familiar categories.

Readable introductions for the non-specialist about the role of *neural networks* (including *recurrent neural networks*) in cognitive processing and neuroscience include P. M. Churchland, *The Engine of Reason, the Seat of the Soul* (Cambridge, MA: MIT Press, 1996) and Manfred Spitzer, *The Mind Within the Net* (Cambridge, MA: MIT Press, 1999). On the role of *chaos* and chaotic processes in the brain, see, e.g. C. Skarda and W. Freeman, "How Do Brains Make Chaos in Order to Make Sense of the World?" *Behavioral and Brain Sciences* (1987) 10: 161–95 and H. Walter, *Neurophilosophy of Free Will* (Cambridge, MA: MIT Press, 2001, Part III). An overview of research on *parallel processing* in visual perception can be found in "Decomposing and Localizing Vision: An Exemplar for Cognitive Neuroscience" by William Bechtel, in *Philosophy and the Neurosciences: A Reader*, ed. by W. Bechtel, Pete Mandik, Jennifer Mundale and Robert Stufflebaum (Oxford: Blackwell Publishers, 2001), pp. 225–49. Introductions to *complex dynamical systems* or *complexity* for non-specialists include R. Lewin, *Complexity: Life at the Edge of Chaos* (New York: Macmillan Publishers, 1992) and M. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Chaos* (New York: Simon and Schuster, 1992). Works that attempt to apply theories about complex dynamical systems to issues about action and agency include E. Thelen and R. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge, MA: MIT Press, 1994) and Alicia Juarrero, *Dynamics in Action: Intentional Behavior as a Complex System* (Cambridge, MA: MIT Press, 1999).