not true (although what it tells us when it is applied to velocities of the magnitudes we usually consider in everyday life comes very, very close to the truth). And yet when one considers this principle in the abstract—in isolation from the considerations that guided Einstein in his development of Special Relativity—it seems to force itself upon the mind as true, to be true beyond all possibility of doubt. It seems, therefore, that the kind of "inner conviction" that sometimes moves one to say things like, "I can just *see* that that proposition *has* to be true" is not infallible. (This is not an isolated example. Consider the case of Euclidean geometry, which seems to force itself upon the mind as the real geometry of the physical world. The physicists tell us, however, that Euclidean geometry is at best *approximately* true of the physical world.)

Nevertheless, a mystery is a mystery. Since compatibilism hides a mystery, should we not therefore be incompatibilists? Unfortunately, incompatibilism also hides a mystery.

Behold, I show you a mystery.

If we are incompatibilists, we must reject either free will or determinism (or both). What happens if we reject determinism? It is a bit easier now to reject determinism than it was in the nineteenth century, when it was commonly believed, and with reason, that determinism was underwritten by physics. But the quantum-mechanical world of current physics is irreversibly indeterministic (at least this is the usual view among physicists), and physics has therefore got out of the business of underwriting determinism. Nevertheless, the physical world is filled with objects and systems that seem to be deterministic "for all practical purposes"—digital computers, for example—and many philosophers and scientists believe that a human organism is deterministic for all practical purposes. But let us not debate this question. Let us suppose for the sake of argument that human organisms display a considerable degree of indeterminism. Let us suppose in fact that each human organism is such that when the human person associated with that organism (we leave aside the question whether the person and the organism are identical) is trying to decide whether to do A or to do B, there is a physically possible future in which the organism behaves in a way appropriate to a decision to do A and that there is also a physically possible future in which the organism behaves in a way appropriate to a decision to do B. We shall see that this supposition leads to a mystery. We shall see that the indeterminism that seems to be required by free will seems also to destroy free will.

Let us look carefully at the consequences of supposing human behavior to be undetermined. Suppose Jane is in an agony of indecision; if her deliberations go one way, she will in a moment speak the words, "John, I lied to you about Alice," and if her deliberations go the other way, she will bite her tongue and remain

silent. We have supposed there to be physically possible continuations of the present in which each of these things happens. Given the whole state of the physical world at the present moment, and given the laws of nature, both these things are possible; either might equally well happen.

Each contemplated action will, of course, have antecedents in Jane's cerebral cortex, for it is in that part of Jane (or of her body) that control over her vocal apparatus resides. Let us make a fanciful assumption about these antecedents, since it will make no real difference to our argument what they are. (It will help us to focus our thoughts if we have some sort of mental picture of what goes on inside Jane at the moment of decision.) Let us suppose that a certain current-pulse is proceeding along one of the neural pathways in Jane's brain and that it is about to come to a fork. And let us suppose that if it goes to the left, she will make her confession, and that if it goes to the right, she will remain silent. And let us suppose that it is undetermined which way the pulse will go when it comes to the fork: even an omniscient being with a complete knowledge of the state of Jane's brain and a complete knowledge of the laws of physics and unlimited powers of calculation could say no more than, "The laws and the present state of her brain would allow the pulse to go either way; consequently, no prediction of what the pulse will do when it comes to the fork is possible; it might go to the left, and it might go to the right, and that's all there is to be said."

Now let us ask: Is it *up to Jane* whether the pulse goes to the left or to the right?[4] If we think about this question for a moment, we shall see that it is very hard to see how this could be up to her. Nothing in the way things are at the instant before the pulse makes its "decision" to go one way or the other makes it happen that the pulse goes one way or goes the other. If it goes to the left, that *just happens*. If it goes to the right, *that* just happens. There is no way for Jane to *influence* the pulse. There is no way for her to make it go one way rather than the other. Or, at least, there is no way for her to make it go one way rather than the other and leave the "choice" it makes an undetermined event. If Jane did something to *make* the pulse go to the left, then, obviously, its going to the left would *not* be an undetermined event. It is a plausible idea that it is up to an agent what the outcome of a process will be only if the agent is able to arrange things in a way that would make the occurrence of *this* outcome inevitable or in a way that would make the occurrence of *that* outcome inevitable. If this plausible idea is right, there would seem to be no possibility of its being up to Jane (or to anyone else) what the outcome of an *indeterministic* process would be. And it seems to follow that if, when one is trying to decide what to do, it is truly undetermined what the outcome of one's deliberations will be, it cannot be up to one what the outcome of one's deliberations will be. It is, therefore, far from clear whether incompatibilism is a tenable position. The incom-

patibilist who believes in free will must say this: it is possible, despite the above argument, for it to be up to an agent what the outcome of an indeterministic process will be. But how is the argument to be met?

Some incompatibilists attempt to meet this argument by means of an appeal to a special sort of causation. Metaphysicians have disagreed about what kinds of things stand in the cause-and-effect relation. This is the orthodox, or "Humean" position: Although our idioms may sometimes suggest otherwise, causes and effects are always events. We may *say* that "Stalin caused" the deaths of millions of people, but when we talk in this way, we are not, in the strictest sense, saying that an *individual thing* (Stalin) was the cause of certain events. It was, strictly speaking, certain *events* (certain actions of Stalin) that were the cause of certain other events (the millions of deaths). It has been suggested, however, that, although events do indeed cause other events, in some cases, *persons* or *agents*, individual things, cause events. According to this suggestion, it might very well be that an event in Jane's brain—a current-pulse taking the left-hand branch of a neural fork, say—had Jane as its cause. And not some event or change involving Jane, not something taking place inside Jane, not something Jane *did*, but Jane herself, the person Jane, the agent Jane, the individual thing Jane.

This "type" of causation is usually labeled 'agent-causation', and it is contrasted with 'event-causation', the other "type" of causation, the kind of causation that occurs when one event causes another event. An event is a change in the intrinsic properties of an individual or a change in the ways certain individuals are related to one another. Event-causation occurs when a change that occurs at a certain time is due to a change that occurred at some earlier time. If there is such a thing as agent-causation, however, some changes are not due to earlier changes but simply to agents: to agents *full stop*; to agents *period*.

Let us now return to the question confronting the incompatibilist who believes in free will: How is it possible for it to be up to an agent what the outcome of an indeterministic process will be? Those incompatibilists who appeal to agent-causation answer this question as follows: "A process's having one outcome rather than one of the other outcomes it might have had is an event. For it to be up to an agent what the outcome of a process will be is for the agent to be able to cause each of the outcomes that process could have. Suppose, for example, that Jane's deciding what to do was an indeterministic process and that this process terminated in her deciding to speak, although, since it was indeterministic, the laws of nature and the way things were when the process was initiated were consistent with its terminating in her remaining silent. But suppose that Jane *caused* the process to terminate in her speaking and that she *once was able* to cause it to terminate in her being silent. Then it was up to her what the outcome was. That is what

it is for it to have been up to an agent whether a process would terminate in A or B: to have caused it to terminate in one of these two ways, and to have been *able* to cause it to terminate in the other."

There are two "standard" objections to this sort of answer. They take the form of questions. The first question is, "But what does one add to the assertion that Jane decided to speak when one says she was the agent-cause of her decision to speak?" The second is, "But what about the event *Jane's becoming the agent-cause of her decision to speak*? According to your position, this event occurred and it was undetermined—for if it were determined by some earlier state of things and the laws of nature, then her decision to speak would have been determined by these same factors. Even if there is such a thing as agent-causation and this event occurred, how could it have been *up to Jane* whether it occurred? And if Jane was the agent-cause of her decision to speak and it was not up to her whether she was the agent-cause of her decision to speak, then it was not up to her whether she would speak or remain silent."

These two standard objections have standard replies. The first reply is, "I don't know how to answer your question. But that is because causation is a mystery, and not because there is any *special* mystery about *agent*-causation. How would *you* answer the corresponding question about event-causation: What does one add to the assertion that two events occurred in succession when one says the earlier was the *cause* of the later?" The second reply is, "But it *was* up to Jane which of the two events *Jane's becoming the agent-cause of her decision to speak* and *Jane's becoming the agent-cause of her decision to remain silent* would occur. This is because she was the agent-cause of the former and was able to have been the agent-cause of the latter. In any case in which Jane is the agent-cause of an event, she is also the agent-cause of her being the agent-cause of that event, and the agent-cause of her being the agent-cause of her being the agent-cause of that event, and so on 'forever.' Of course, she is no doubt not *aware* of being the agent-cause of all these events, but the doctrine of agent-causation does not entail that agents are aware of all the events of which they are agent-causes."

Perhaps these replies are effective and perhaps not. I reproduce them because they are, as I have said, standard replies to standard objections. I have no clear sense of what is going on in this debate because I do not understand agent-causation. At least I don't think I understand it. To me, the suggestion that an individual thing, as opposed to a *change* in an individual thing, could be the cause of a change is a mystery. I do not intend this as an argument against the *existence* of agent-causation—of some relation between individual things and events that, when it is finally comprehended, will be seen to satisfy the descriptions of "agent-causation" that have been advanced by those who claim to grasp this concept. The world is full of mysteries. And there are many phrases that seem to some to be

nonsense but which are in fact not nonsense at all. ("Curved space! What nonsense! Space is what things that are curved are curved *in*. Space itself can't be curved." And no doubt the phrase 'curved space' *wouldn't* mean anything in particular if it had been made up by, say, a science-fiction writer and had no actual use in science. But the general theory of relativity does imply that it is possible for space to have a feature for which, as it turns out, those who understand the theory all regard 'curved' as an appropriate label.) I am saying only that agent-causation is a mystery and that to explain, by an appeal to agent-causation, how it could be up to someone what the outcome of an indeterministic process would be, is to explain a mystery by a mystery.

But now a disquieting possibility suggests itself. Perhaps the explanation of the fact that both compatibilism and incompatibilism seem to lead to mysteries is simply that the concept of free will is self-contradictory. Perhaps free will is, as the incompatibilists say, incompatible with determinism. But perhaps it is also incompatible with *in*determinism, owing to the impossibility of its being up to an agent what the outcome of an indeterministic process will be. If free will is incompatible with both determinism and indeterminism, then, since either determinism or indeterminism has to be true, free will is impossible. And, of course, what is impossible does not exist. Can we avoid mystery by accepting the non-existence of free will? If we are willing to say that free will does not exist, then we need not reject the Principle—*and* we need not suppose it is possible for it to be up to an agent what the outcome of an indeterministic process will be.

But consider. Suppose you are trying to decide what to do. And suppose the choice that confronts you is not a trivial one. Let us not suppose you are trying to decide which of two movies to see or which flavor of ice cream to order. Let us suppose the matter to be one of great importance—of great importance to *you*, at any rate. You are, perhaps, trying to decide whether to marry a certain person or whether to risk losing your job by reporting unethical conduct on the part of a superior or whether to sign a "do not resuscitate" order on behalf of a beloved relative who is critically ill. Pick one of these situations and imagine you are in it. (If you are in fact faced with a non-trivial choice, you have no need to imagine anything. Think of your own situation.) Consider the two contemplated courses of action. Hold them before your mind's eye, and let your attention pass back and forth between them. Do you really think it isn't up to you which of these courses of action you will choose? Can you really believe that?

Many philosophers have said that although the choice between contemplated *future* courses of action always seems "open" to them at the time, when they look back on their *past* decisions, the particular decision they have made always or almost always seems inevitable from that perspective. Is this a plausible thesis? I can testify that I do not myself find any such thing when I examine my past decisions.

And, even if I did, I should regard it as an open question whether "foresight" or "hindsight" was more to be trusted. (Why should we suppose that hindsight is trustworthy? Maybe there is within us some psychological mechanism that produces the illusion of the inevitability of our past decisions in order to enable us more effectively to put these decisions behind us and to spare us endless retrospective agonizing over them. Maybe we have a natural tendency to interpret our past decisions in a way that presents them in the best possible light. One can think of lots of not implausible hypotheses that imply that our present impression that our past decisions were the only possible ones—if we indeed have this impression—is untrustworthy.)

When I myself look at contemplated future courses of action in the way I have described above, I discover an irresistible tendency to believe that each of them is "open" to me. This tendency may be a vehicle of illusion. It may be that free will belongs to appearance, not to reality. If the concept of free choice were self-contradictory, a belief in this self-contradictory thing might nevertheless be indispensable to human action. Let us ask ourselves: "What would it be like to believe, really to *believe*, that only one course of action is ever open to me?"

It can plausibly be argued that it would be impossible under such circumstances ever to try to decide what to do. Suppose, for example, that you are in a certain room, a room with a single door, and that this door is the only possible way out of the room. Suppose that, as you are thinking about whether to leave the room, you hear a click that may or may not have been the sound of the door's being locked. You are now in a state of uncertainty about whether the door is locked and are therefore in a state of uncertainty about whether it is possible for you to leave the room. Can you continue to try to decide whether to leave the room? It would seem not. (Try the experiment of imagining yourself in this situation and seeing whether you can imagine yourself continuing to try to decide whether to leave.) You cannot because you no longer believe it's possible for you to leave the room. (It's not that you believe it's *im*possible for you to leave the room. You don't believe that either, for you are in a state of uncertainty about whether it is possible for you to leave.) You can, of course, try to decide whether to get up and try the door. But that *is*—at least you probably believe this—possible for you. And you can try to decide, "conditionally," whether to leave the room *if* the door should prove to be unlocked. But that is not the same thing as trying to decide whether to leave the room.

This thought-experiment convinces me that I cannot try to decide whether to do A or B unless I believe that doing A and doing B are both possible for me. And, therefore, I am convinced that I could not try to decide what to do unless I believed that more than one course of action was sometimes open to me. And if I

never tried to decide what to do, if I never deliberated, I should not be a very effective human being. In the state of nature, I should no doubt starve. In a civilized society, I should probably have to be institutionalized. Belief in one's own free will is therefore something we can hardly do without. It would seem to be an evolutionary necessity that beings like ourselves should believe in their own free will. And evolutionary necessity has scant respect for such niceties as logical consistency. It is arguable, therefore, that we cannot trust our conviction that we have free will (if, indeed, we do have this conviction). If evolution has forced a certain belief on us (for the simple reason that we can't survive without that belief), the fact that we hold it provides no evidential support for the hypothesis that the belief is true; it does not even support the hypothesis that that belief is logically consistent. (*Aren't* there people who think that no one, themselves included, has free will? Well, there are certainly people who *say* they think this. I suspect they are not describing their own beliefs correctly. But even if there are people who think no one has free will, it does not follow that these people do not think they have free will, for people do have contradictory beliefs. It may be that "on one level"—the abstract and theoretical—certain people believe free will to be an illusion, while on another level—the concrete and everyday—they believe themselves to have free will.)

Nevertheless, when all is said and done, I find myself with the belief that sometimes more than one course of action is open to me, and I cannot give it up. (As Dr. Johnson said, "Sir, we know our will is free, and there's an end on't.") And I don't find the least plausibility in the hypothesis that this belief is illusory. It can sometimes seem attractive to think of free will as an illusion. To think of free will as an illusion—or to toy with the idea in a theoretical sort of way—can be attractive to someone who has betrayed a friend or achieved success by spreading vicious rumors. If you had done something of that sort, wouldn't you want to believe that you couldn't have done otherwise, that no other course of action was really open to you? Wouldn't it be tempting to suppose that your actions were determined by your genes and your upbringing or by the way things were thousands or millions of years ago? (Jean-Paul Sartre once remarked that determinism was a bottomless well of excuses.) And it is immensely attractive to suppose oneself to be a member of an intellectual élite whose members have freed themselves from an illusion to which the mass of humanity is subject. The hypothesis has its unattractive aspects too, of course. For one thing, if it rules out blame, it presumably rules out praise on the same grounds. But, however attractive or unattractive it may be, it just seems to be false. If some unimpeachable source—God, say—were to tell me I didn't have free will, I'd have to regard that piece of information as proof that I didn't understand the World at all. It would be as if an unimpeachable source had told me that

consciousness did not exist or that the physical world was an illusion or that self-contradictory statements could be true. I'd have to say, "Well, all right. You *are* an unimpeachable source. But I just don't see how what you're telling me could be true." In short: to propose that we believe that we do not have free will is to propose that we accept a mystery.

I conclude that there is no position one can take concerning free will that does not confront its adherents with mystery. I myself prefer the following mystery: I believe that the outcome of our deliberations about what to do is undetermined and that it is nevertheless—in some way I have no shadow of an understanding of—sometimes up to us what the outcome of these deliberations will be.

I believe that if Jane has freely decided to speak then the following must be true: if God were to create a thousand perfect duplicates of Jane as she was an instant before the decision to speak was made and were to place each one in circumstances that perfectly duplicated Jane's circumstances at that instant, some of the duplicates would choose to speak and some of them would choose to remain silent, and there would be no explanation whatever of the fact that any particular duplicate made whichever choice it was she made. And yet, I believe, Jane, the one actual Jane, was able to speak and able to remain silent.

I accept this mystery because it seems to me to be the smallest mystery available. If someone believes that human beings do not have free will, that person accepts a mystery—and, in my view, a greater, deeper mystery than the one I accept. Someone who denies the Principle, for example, accepts a mystery—and, in my view, a greater, deeper mystery than the one I accept. But others may judge the sizes of these mysteries differently.

It is important to be aware that we have not said everything there is to say about the size of the mysteries connected with the free-will problem. The most important of the topics we have not discussed in this connection is the relation between free will and morality. In our preliminary discussion of the concept of free will, we said it was a common opinion that free will was required by morality. If this common opinion is correct, then, in a world without free will, all moral judgments are false or in some other way out of place. If that were so, it would greatly aggravate the mystery confronting those who deny the existence of free will. Could it really be, for example, that racism or child abuse or genocide or serial murder are morally unobjectionable? If an unimpeachable source were to inform me that child abuse was morally unobjectionable, my dominant reaction would be one of horror. But I should also have a negative reaction to this revelation that was more intellectual, more theoretical. I should have to conclude that I didn't understand the World at all. I should have to say I simply didn't understand how it could *be* that there was nothing morally objectionable about child abuse.

It is, however, controversial whether a philosopher who rejects free will must concede that all moral judgments are false (or are all in some other way vehicles of illusion). The "common opinion" that morality requires free will is not so common as it used to be. When almost all English-speaking philosophers were compatibilists, this opinion was held by almost everyone in the English-speaking philosophical world. It was the common assumption of the compatibilists and the few incompatibilists there were. Now, however, compatibilists are less common than they used to be, owing principally to the fact that philosophers have come to realize that a compatibilist must reject the Principle. Many philosophers now reject compatibilism who might previously have been strongly attracted to this position. And because these philosophers, or many of them, reject the possibility of any sort of free will that requires indeterminism, they reject free will altogether. But most philosophers who reject free will are not willing to say that morality is an illusion. It has, therefore, become an increasingly popular position that morality does not require free will after all. For this reason, I have not included the thesis that morality is an illusion among the mysteries that must be accepted by those who reject free will. I myself continue to believe that morality is an illusion if there is no free will. (In fact, this conditional statement seems self-evident to me; if an unimpeachable source told me it was false, I'd regard its falsity as a great mystery.) But, since the issues involved in the debate about this thesis pertain to moral philosophy and not to metaphysics, I will not discuss them.

However one may judge the relative "sizes" of the mysteries that confront the adherents of the various positions one might take concerning free will, these mysteries exist. The metaphysician's task is to display these mysteries. Each of us must decide, with no further help from the metaphysician, how to respond to the array of mysteries that the metaphysician has placed before us.

Suggestions for Further Reading

Berofsky's *Free Will and Determinism* and Watson's *Free Will* are excellent collections devoted to the problem of free will and determinism. Fischer's more recent *Moral Responsibility* contains much useful material. My own book, *An Essay on Free Will*, is a defense of incompatibilism. Large parts of it are accessible to those without formal philosophical training. The central argument of the book is attacked in Lewis's superb article, "Are We Free to Break the Laws?" (rather difficult for those without philosophical training). Dennett's *Elbow Room* is a highly readable (if somewhat idiosyncratic) defense of compatibilism. The question, 'Could there be free will in an indeterministic world?' is the main topic of the essays in O'Connor's *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. For