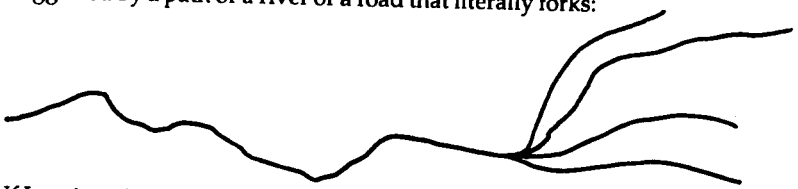
11

The Powers of Rational Beings: Freedom of the Will

WE NOW TURN TO another mystery, a mystery about the *powers* of rational beings; that is, a mystery about what human beings are able to do. This mystery is the mystery of free will and determinism. The best way to get an intuitive grip on the problem of free will and determinism is to think of time as a "garden of forking paths." That is, to think of the alternatives that one considers when one is deciding what to do as being parts of various "alternative futures" and to think of these alternative futures diagrammatically, in the way suggested by a path or a river or a road that literally forks:

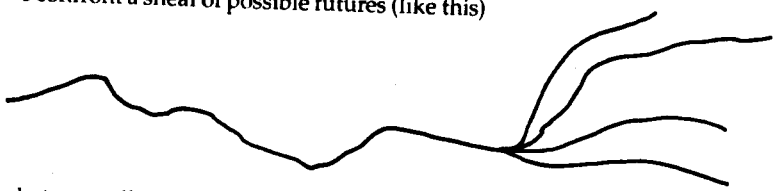


If Jane is trying to decide whether to tell all or to continue her life of deception, she is in a situation strongly analogous to that of someone who is hesitating between forks in a road. That is why this sort of diagram is so suggestive. Let us apply this idea to the problem of free will and determinism.

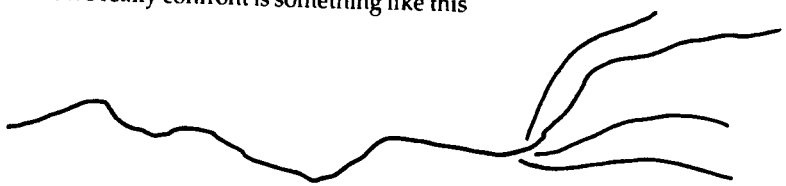
To say that one has free will is to say that when one decides among forks in the road of time (or, more prosaically, when one decides what to do), one is at least sometimes able to take more than one of the forks. Thus, Jane, who is deciding between a fork that leads to telling all and a fork that leads to a life of continued deception, has free will (on this particular occasion) if she is able to tell all and is also able to continue living a life of deception. One has free will if sometimes more than one of the forks in the road of time is "open" to one. One lacks free will if on every occasion on which one must make a decision only one of the forks before one—of course it will be the fork one in fact takes—is open to one. If John is locked in a room and doesn't know that he is locked in, and if he is in the process of deliberating about whether to leave, one of the alternative futures he is contemplating—leaving—is, in point of fact, not open to him, and he thus lacks free will in the matter of staying or leaving.[1]

It is a common opinion that free will is required by morality. Let us examine this common opinion from the perspective that is provided by looking at time as a garden of forking paths. While it is obviously false—for about six independent reasons—that the whole of morality consists in making judgments of the form 'You should not have done X', we can at least illustrate certain important features of the relation between free will and morality by examining the relation between the concept of free will and the content of such judgments. The judgment that you shouldn't have done X implies that you should have done something else instead; that you should have done something else instead implies that there was something else for you to do; that there was something else for you to do implies that you *could* have done something else; that you could have done something else implies that you have free will. To make a moral judgment about one of your acts is to evaluate your taking one of the forks in the road of time, to characterize it as better or worse than various of the other forks that were open to you. (Note that if you have made a choice by taking one of the forks in what is literally a road, no one could blame you for taking the fork you did if all of the other forks were blocked.) A moral evaluation of what someone has done requires two or more alternative possibilities of action for that person just as surely as a contest requires two or more contestants.

Let us now see what help the conception of time as a garden of forking paths gives us in understanding what is meant by determinism. Determinism is the thesis that it is true at every moment that the way things then are determines a unique future, that only *one* of the alternative futures that may exist relative to a given moment is a physically possible continuation of the state of things at that moment. Or, if you like, we may say that determinism is the thesis that only one continuation of the state of things at a given moment is consistent with the laws of nature. (For it is the laws of nature that determine what is physically possible. It is, for example, now physically possible for you to be in Chicago at noon tomorrow if and only if your being in Chicago at noon tomorrow is consistent with both the present state of things and the laws of nature.) Thus, according to determinism, although it may often seem to us that we confront a sheaf of possible futures (like this)



what we really confront is something like this



This figure is almost shaped like a road that splits into four roads, but not quite: three of the four "branches" that lead away from the "fork" are not connected with the original road, although they come very close to it. (Thus they are not really branches in the road, and the place at which they almost touch the road is not really a fork.) If we were to view this figure from a distance—across the room, say—it would seem to us to have the shape of a road that forks. We have to look at it closely to see that what appeared from a distance to be three "branches" are not connected with the long line or with one another. In the figure, the point at which the three unconnected lines *almost* touch the long line represents the present. The unconnected lines represent futures that are not physically possible continuations of the present, and the part of the long line to the right of the "present" represents a future that is a physically possible continuation of the present. The gaps between the long line and the unconnected lines represent causal discontinuities, violations of the laws of nature—in a word, miracles. The reason these futures are not physically possible continuations of the present is that "getting into" any of them from the present would require a miracle. The fact that the part of the long line that lies to the right of the "present" actually proceeds from that point represents the fact that this line-segment corresponds to a physically possible future.

This figure, then, represents four futures, three of which are physically impossible and exactly one of which is physically possible. If these four futures are the only futures that "follow" the present, then this figure represents the way in which each moment of time must be if the universe is deterministic: each moment must be followed by exactly one physically possible future.

The earlier diagram, however, represents an indeterministic situation. The road really does fork. The present is followed by four possible futures. Any one of them could, consistently with the laws of nature, evolve out of the present. Any one of them could, consistently with the laws of nature, turn out to be the actual future. Therefore, it is only if the universe is indeterministic that time *really is* a "garden of forking paths." But even in a deterministic universe, time could *look like* a garden of forking paths. Remember that our figure, when viewed from across the room, *looked* as if it had the shape of a road that forked. We cannot see all, or even very many, of the causes that operate in any situation. It could be, therefore, that the universe is deterministic, even though it looks to our limited vision as if there were sometimes more than one possible future. It may look to Jane as if she faces two possible futures, in one of which she tells all and in the other of which she continues her life of deception. But it may well be that the possibility of one or the other of these contemplated futures is mere appearance—an illusion, in fact. It may be that, in reality, causes already at work in her brain and central nervous system and immediate environment have already "ruled out" one or the other of these futures: it may be that one or the other of them is such that it could not come to pass unless a physically impossible event, a miracle, were to happen in her brain or central nervous system or environment.

Ask yourself this question. What would happen if some supernatural agency—God, say—were to "roll history back" to some point in the past and then "let things go forward again"? Suppose the agency were to cause things

to be once more just as they were at high noon, Greenwich time, on 11 March 1893 and were thereafter to let things go on of their own accord. Would history literally repeat itself? Would there be two world wars, each the same in every detail as the wars that occurred the "first time around"? Would a president of the United States called 'John F. Kennedy' be assassinated in Dallas on the date that on the new reckoning is called '22 November 1963'? Would you, or at least someone exactly like you, exist? If the answer to these questions is No, then determinism is false. Equivalently, if determinism is true, the answer to these questions is Yes. If determinism is true, then, if the universe were rolled back to a previous state by a miracle, and if there were no further miracles, the history of the world would repeat itself. And if the universe were rolled back to a previous state thousands of times, this exact duplication would happen every time. If there are no forks in the road of time—if all of the apparent forks are merely apparent, illusions due to our limited knowledge of the causes of things—then restoring the universe to some earlier condition is like moving a traveler on a road without forks back to an earlier point on that road. If there are no forks in the road, then, obviously enough, the traveler must traverse the same path a second time.

It has seemed obvious to most people who have not been exposed (perhaps 'subjected' would be a better word) to philosophy that free will and determinism are incompatible. It is almost impossible to get beginning students of philosophy to take seriously the idea that there could be such a thing as free will in a deterministic universe. Indeed, people who have not been exposed to philosophy usually understand the word 'determinism' (if they know the word at all) to stand for the thesis that there is no free will. And you might think that the incompatibility of free will and determinism deserves to seem obvious—because it is obvious. To say that we have free will is to say that more than one future is sometimes open to us. To affirm determinism is to say that every future that confronts us but one is physically impossible. And, surely, a physically impossible future can't be open to anyone, can it? If we know that a "Star Trek" sort of future is physically impossible (because, say, the "warp drives" and "transporter beams" that figure essentially in such futures are physically impossible), then we know that a "Star Trek" future is not open to us or to our descendants.

People who are convinced by this sort of reasoning are called *incompatibilists*: they hold that free will and determinism are incompatible. As I have hinted, however, many philosophers are *compatibilists*: they hold that free will and determinism are compatible. Compatibilism has an illustrious history among English-speaking philosophers, a history that embraces such figures as the seventeenth-century English philosopher Thomas Hobbes, the eighteenth-century Scottish philosopher David Hume, and the nineteenth-century English philosopher John Stuart Mill. And the majority of twentieth-century English-speaking philosophers have been compatibilists. (But compatibilism has not had many adherents on the continent of Europe. Kant, for example, called it a "wretched subterfuge.")

A modern compatibilist can be expected to reply to the line of reasoning I have just presented in some such way as follows:

Yes, a future, in order to be open to one, does need to be physically possible. It can't, for example, contain faster-than-light travel if faster-than-light travel is physically impossible. But we must distinguish between a future's being physically possible and its having a physically possible connection with the present. A future is physically possible if everything that happens in it is permitted by the laws of nature. A future has a physically possible connection with the present if it could be 'joined' to the present without any violation of the laws of nature. A physically possible future that does not have a physically possible connection with the present is one that, given the present state of things, would have to be 'inaugurated' by a miracle, an event that violated the laws of nature, but in which, thereafter, events proceeded in accordance with the laws. Determinism indeed says that of all the physically possible futures, one and only one has a physically possible connection with the present—one and only one could be joined to the present without a violation of the laws of nature. My position is that some futures that could not be joined to the present without a violation of the laws of nature are, nevertheless, open to us.

Two philosophical problems face the defenders of compatibilism. The easier is to provide a clear statement of *which* futures that do not have a physically possible connection with the present are "open" to us. The more difficult is to make it seem at least plausible that futures that are in this sense open to an agent really deserve to be so described.

An example of a solution to these problems may make the nature of the problems clearer. The solution I shall briefly describe would almost certainly be regarded by all present-day compatibilists as defective, although it has a respectable history. I choose it not to suggest that compatibilists can't do better but simply because it can be described in fairly simple terms.

According to this solution, a future is open to an agent, if, given that the agent chose that future (chose that path leading away from a fork in the road of time), it would come to pass. Thus it is open to me to stop writing this book and do a little dance because, if I so chose, that's what I'd do. But if Alice is locked in a prison cell, it is not open to her to leave: if she chose to leave, her choice would be ineffective because she would come up against a locked prison door. Now consider the future I said was open to me—to stop writing and do a little dance—and suppose that determinism is true. Although a choice on my part to behave in that remarkable fashion would (no doubt) be effective if it occurred, it is as a matter of fact *not* going to occur, and, therefore, given determinism, it is determined by the present state of things and the laws of nature that such a choice is not going to occur. It is in fact determined that *nothing* is going to occur that would have the consequence that I stop writing and do a little dance. Therefore, none of the futures in which I act in that bizarre way is a future that has a physically possible connection with the present: such a future could come to pass only if it were inaugurated by an event of a sort that is ruled out by the present state of things and the laws of nature.

And yet, as we have seen, many of these futures are "open" to me in the sense of 'open' that the compatibilist has proposed.

Is this a reasonable sense to give to this word? (We now take up the second problem that confronts the compatibilist.) This is a very large question. The core of the compatibilist's answer is an attempt to show that the reason we are interested in open or accessible futures is that we are interested in modifying the way people behave. One important way in which we modify behavior is by rewarding behavior that we like and punishing behavior that we dislike. We tell people that we will put them in jail if they steal and that they will get a tax break if they invest their money in such-and-such a way. But there is no point in trying to get people to act in a certain way if that way is not in some sense open to them. There is no point in telling Alfred that he will go to jail if he steals unless it is somehow open to him not to steal.

And what is the relevant sense of "open"? Just the one I have proposed, says the compatibilist. One modifies behavior by modifying the choices people make. That procedure is effective just insofar as choices are effective in producing behavior. If Alfred chooses not to steal (and remains constant in that choice), then he won't steal. But if Alfred chooses not to be subject to the force of gravity, he will nevertheless be subject to the force of gravity. Although it would no doubt be socially useful if there were some people who were not subject to the force of gravity, there is no point in threatening people with grave consequences if they do not break the bonds of gravity, for even if you managed to induce some people to choose not to be subject to the force of gravity, their choice would not be effective. Therefore (the compatibilist concludes), it is entirely appropriate to speak of a future as "open" if it is a future that would be brought about by a choice—even if it were a choice that was determined not to occur. And if Alfred protests when you punish him for not choosing a future that was in this sense open to him, on the ground that it was determined by events that occurred before his birth that he not make the choice that would have inaugurated that future—if he protests that only a *miracle* could have inaugurated such a future—you can tell him that his punishment will not be less effective in modifying his behavior (and the behavior of those who witness his punishment) on *that* account.

When things are put that way, compatibilism can look like nothing more than robust common sense. Why, then, do people have so much trouble believing it? Why does it arouse so much resistance? I think that the reason is that compatibilists can make their doctrine seem like robust common sense only by sweeping a mystery under the carpet and that, despite their best efforts, the bulge shows. People are aware that something is amiss with compatibilism even when they are unable to articulate their misgivings. I believe that it is possible to lift the carpet and display the hidden mystery. The notion of "not having a choice" has a certain logic to it. One of the principles of this logic is, or so it seems, embodied in the following thesis, which I shall refer to as the No Choice Principle:

Suppose that p and that no one has (or ever had) any choice about whether p. And suppose also that the following conditional (if-then) statement is

> true and that no one has (or ever had) any choice about whether it is true: if p, then q. It follows from these two suppositions that q and that no one has (or ever had) any choice about whether q.

In this statement of the No Choice Principle, any declarative sentences can replace the symbols 'p' and 'q'. (But the same sentence must replace 'p' at each place it occurs, and the same goes for 'q'.) We might, for example, replace 'p' with 'Plato died long before I was born' and 'q' with 'I have never met Plato':

> Suppose that Plato died long before I was born and that no one has (or ever had) any choice about whether Plato died long before I was born. And suppose also that the following conditional statement is true and that no one has (or ever had) any choice about whether it is true: if Plato died long before I was born, then I have never met Plato. It follows from these two suppositions that I have never met Plato and that no one has (or ever had) any choice about whether I have never met Plato.

The No Choice Principle seems undeniably correct. How could I have a choice about anything that is an inevitable consequence of something I have no choice about? And yet, as we shall see, the compatibilist must deny the No Choice Principle. To see why this is so, let us suppose that determinism is true and that the No Choice Principle is correct. Now let us consider some state of affairs that we should normally suppose someone had a choice about. Consider, say, the fact that I am writing this book. Most people—at least most people who knew I was writing a book—would assume that I had a choice about whether I was engaged in this project. They would assume that it was open to me to have undertaken some other project or no project at all. But we are supposing that determinism is true, and that means that ten million years ago (say) there was only one physically possible future, a future that included my being engaged in writing this book at the present date (since that is what I am in fact doing): given the way things were ten million years ago and given the laws of nature, it had to be true that I was now engaged in writing this book. But consider the two statements

- Things were thus-and-so ten million years ago.
- If things were thus-and-so ten million years ago, then I am working on this book now.

(Here 'thus-and-so' is a sort of gesture at a complete description or specification of the way things were ten million years ago.) Each of these statements is true. And it is obvious that no one has or ever had any choice about the truth of either. It is obvious that no one—no human being, certainly—has or ever had any choice about whether things *were* thus-and-so ten million years ago, since at that time the first human beings were still millions of years in the future.

And no one has any choice about whether the second statement, the if-then statement, is true because this statement is a consequence of the laws of nature, and no one—no human being, certainly—has any choice about what the laws of nature are. If we imagine a possible world in which, as in the actual world, things were thus-and-so ten million years ago, and in which, unlike in the actual world, I decided to learn to sail instead of writing this book, we are imagining a world in which the laws of nature are different; for the *actual* laws dictate that if at some point in time things are thus-and-so, then, ten million years later I (or at any rate someone just like me) shall be writing and not sailing.

But if both of the above statements are true, then it follows, by the No Choice Principle, that neither I nor anyone else has or ever had any choice about whether I write this book. And, obviously, the content of the particular example—my writing a book—played no role in the derivation of this conclusion. It follows that, given the No Choice Principle, determinism implies that there is no free will. That is why the compatibilist must reject the No Choice Principle. This is the hidden mystery that, I contend, lies behind the façade of bluff common sense that compatibilism presents to the world: the compatibilist must reject the No Choice Principle, and the No Choice Principle seems to be true beyond all possibility of dispute. (Either that or the compatibilist must hold that one can have a choice about what went on in the world before there were any human beings or that one can have a choice about what the laws of nature are. But these alternatives look even more implausible than a rejection of the No Choice Principle.) If the No Choice Principle were false, that would be a great mystery indeed.

We must not forget, however, that mysteries really do exist. There are principles that are commonly held, and with good reason, to be false and whose falsity seems to be just as great a mystery as the falsity of the No Choice Principle would be. Consider, for example the principle that is usually called "the Galilean Law of the Addition of Velocities." This principle is a generalization of cases like the following. Suppose that an airplane is flying at a speed of 800 kilometers per hour relative to the ground; suppose that inside the aircraft a housefly is buzzing along at a speed of 30 kilometers per hour relative to the airplane in the direction of the airplane's travel; then the fly's speed relative to the ground is the sum of these two speeds: 830 kilometers per hour. According to the Special Theory of Relativity, an immensely useful and well-confirmed theory, the Galilean Law of the Addition of Velocities does not hold (although it comes very, very close to holding when it is applied to velocities of the magnitude that we usually consider in everyday life). 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.

Nevertheless, a mystery is a mystery. If compatibilism hides a mystery, should we therefore be incompatibilists? Unfortunately, incompatibilism also hides a mystery. Behold, I will show you a mystery.

If we are incompatibilists, we must reject either free will or determinism. 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 be-

lieved, and with reason, that determinism was underwritten by physics. But the quantum-mechanical world of current physics seems to be irreversibly indeterministic, 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 that human behavior is undetermined. Suppose that 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 lip and remain silent. We have supposed that there is a physically possible future 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 of 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 there is a certain current-pulse that 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: Does Jane have any choice about whether the pulse goes to the left or to the right? If we think about this question for a moment, we shall see that it is very hard to see how she could have any choice about that. 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 the only way to have a choice about the outcome of a process is to be able to arrange things in ways that will make it inevitable that this or that outcome occur. If this plausible idea is right, then it would seem that there is no way in which anyone could have any choice about the outcome of an indeterministic process. 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, then one could have no choice about that outcome. It is, therefore, far from clear that incompatibilism is a tenable position. The incompatibilist who believes in free will must say this: it is possible, despite the above argument, for one to have a choice about the outcome of an indeterministic process. But how is the argument to be met?