the circumstances, internal and external, motives, perceptions, dispositions, impressions, passions, inclinations taken together, I am still in a contingent state, or whether I am necessitated to make the choice, for example, to go out; that is to say, whether this proposition true and determined in fact, *In all these circumstances taken together I shall choose to go out,* is contingent or necessary.'[11] Leibniz's answer might be put as follows: in one sense of the terms 'necessary' and 'contingent', the proposition 'In all these circumstances taken together I shall choose to go out', may be said to be contingent and not necessary, and in another sense of these terms, it may be said to be necessary and not contingent. But the sense in which the proposition may be said to be contingent, according to Leibniz, is only this: there is no logical contradiction involved in denying the proposition. And the sense in which it may be said to be necessary is this: since 'nothing ever occurs without cause or determining reason', the proposition is causally necessary. 'Whenever all the circumstances taken together are such that the balance of deliberation is heavier on one side than on the other, it is certain and infallible that that is the side that is going to win out'. But if what we have been saying is true, the proposition 'In all these circumstances taken together I shall choose to go out', may be causally as well as logically contingent. Hence we must find another interpretation for Leibniz's statement that our motives and desires may incline us, or influence us, to choose without thereby necessitating us to choose.

Let us consider a public official who has some moral scruples but who also, as one says, could be had. Because of the scruples that he does have, he would never take any positive steps to receive a bribe—he would not actively solicit one. But his morality has its limits and he is also such that, if we were to confront him with a *fait accompli* or to let him see what is about to happen ($10,000 in cash is being deposited behind the garage), then he would succumb and be unable to resist. The general situation is a familiar one and this is one reason that people pray to be delivered from temptation. (It also justifies Kant's remark: 'And how many there are who may have led a long blameless life, who are only *fortunate* in having escaped so many temptations'.)[12] Our relation to the misdeed that we contemplate may not be a matter simply of being able to bring it about or not to bring it about. As St. Anselm noted, there are at least four possibilities. We may illustrate them by reference to our public official and the event which is his receiving the bribe, in the following way: (i) he may be able to bring the event about himself (*facere esse*), in which case he would actively cause himself to receive the bribe; (ii) he may be able to refrain from bringing it about himself (*non facere esse*), in which case he would not himself do anything to insure that he receive the bribe; (iii) he may be able to do something to prevent the event from occurring (*facere non esse*), in which case he would make sure that the $10,000 was *not* left behind the garage; or (iv) he may be unable to do anything to prevent the event from occurring (*non facere non esse*), in which case, though he may not solicit the bribe, he would allow himself to keep it.[13] We have envisaged our official as a man who can resist the temptation to (i) but cannot resist the temptation to (iv): he can refrain from bringing the event about himself, but he cannot bring himself to do anything to prevent it.

Let us think of 'inclination without necessitation', then, in such terms as these. First we may contrast the two propositions:

(1) He can resist the temptation to do something in order to make A happen;
(2) He can resist the temptation to allow A to happen (i.e. to do nothing to prevent A from happening).

We may suppose that the man has some desire to have A happen and thus has a motive for making A happen. His motive for making A happen, I suggest, is one that *necessitates* provided that, because of the motive, (1) is false; he cannot resist the temptation to do something in order to make A happen. His motive for making A happen is one that *inclines* provided that, because of the motive, (2) is false; like our public official, he cannot bring himself to do anything to prevent A from happening. And therefore we can say that this motive for making A happen is one that *inclines but does not necessitate* provided that, because of the motive, (1) is true and (2) is false; he can resist the temptation to make it happen but he cannot resist the temptation to allow it to happen.

NOTES

1. The general position to be presented here is suggested in the following writings, among others: Aristotle, I *Eudemian Ethics,* bk. ii ch. 6, *Nicomachean Ethics,* bk. iii, ch. 1–5; Thomas Reid, *Essays on the Active Powers of Man.* C. A. Campbell, 'Is "Free Will" a Pseudo-Problem?' *Mind,* 1951, 441–65, Roderick M. Chisholm, 'Responsibility and Avoidability', and Richard Taylor, 'Determination and the Theory of Agency', in *Determinism and Freedom in the Age of Modern Science,* ed. Sidney Hook (New York, 1958).
2. Thomas Reid, *Essays on the Active Powers of Man,* essay iv. ch. 4 (*Works,* 600).
3. *Summa Theologica,* First Part of the Second Part, qu. vi ('On the Voluntary and Involuntary').
4. Jonathan Edwards, *Freedom of the Will* (New Haven, 1957): G. E. Moore, *Ethics* (Home University Library, 1912). ch. 6.
5. A. I. Melden, *Free Action* (London, 1961), especially ch. 3. Mr. Melden's own views, however, are quite the contrary of those that are proposed here.
6. Aristotle, *Physics,* bk. iii. ch. 3; Suarez, *Disputations Metaphysicae,* Disputation 18, s. 10.
7. Reid, *Works.* 524.
8. *Essay concerning Human Understanding,* bk. ii, ch. 21.
9. Melden, 166.
10. Reid, *Works,* 608, 612.
11. 'Lettre à Mr. Coste de la Nécessité et de la Contingence' (1707) in *Opera Philosophica,* ed. Erdmann, 447–9.
12. In the Preface to the *Metaphysical Elements of Ethics,* in *Kant's Critique of Practical Reason and Other Works on the Theory of Ethics,* ed. T. K. Abbott (London, 1959), 303.
13. Cf. D. P. Henry, 'Saint Anselm's *De "Grammatico"*', Philosophical Quarterly, x (1960), 115–26. St. Anselm noted that (i) and (iii), respectively, may be thought of as forming the upper left and the upper right corners of a square of opposition, and (ii) and (iv) the lower left and the lower right.

From: Reason & Responsibility, ed by Joel Feinberg and Russ Shafer-Landau, 11th edition (Wadsworth, 2002)

Free Will: Ancient Dispute, New Themes*

ROBERT KANE

Robert Kane teaches philosophy at the University of Texas.

I

"THERE IS A DISPUTATION that will continue till mankind are raised from the dead, between the necessitarians and the partisans of free will." These are the words of twelfth-centnry Persian poet, Jalalu'ddin Rumi. The problem of free will and necessity (or determinism), of which Rumi speaks, has puzzled the greatest minds for centuries—including famous philosophers, literary figures, theologians, scientists, legal theorists, and psychol-

13

ogists—as well as many ordinary people. It has affected and been affected by both religion and science.

In his classic poem, *Paradise Lost,* John Milton describes the angels debating how some of them could have sinned of their own free wills given that God had made them intelligent and happy.[1] Why would they have done it? And why were they responsible for it rather than God, since God had made them the way they were and had complete foreknowledge of what they would do? While puzzling over such questions, even the angels, Milton tells us, were "in Endless Mazes lost" (not a comforting thought for us humans). On the scientific front, issues about free will lead us to ask about the nature of the physical universe and our place in it (are we determined by physical laws and movements of the atoms?), about human psychology and the springs of action (can our actions be predicted by those who know our psychology?), about social conditioning, moral responsibility, crime and punishment, right and wrong, good and evil, and much more.

To dive into these questions, the best way to begin is with the idea of *freedom* itself. Nothing could be more important than freedom to the modern world. All over the globe, the trend (often against resistance) is toward societies that are more free. But why do we want freedom? The simple, and not totally adequate, answer is that to be more free is to have the capacity and opportunity to satisfy more of our desires. In a free society we can walk into a store and buy almost anything we want. We can choose what movies to see, what music to listen to, whom to vote for.

But these are what you might call *surface* freedoms. What is meant by *free will* runs deeper than these everyday freedoms. To see how, suppose we had maximal freedom to make such choices to satisfy our desires and yet the choices we actually made were manipulated by others, by the powers-that-be. In such a world we would have a great deal of everyday freedom to do whatever we wanted, yet our free will would be severely limited. We would be free to act or choose as we will, but would not have the ultimate say about what it is that we will. Someone else would be pulling the strings, not by coercing us against our wishes, but by manipulating us into having the wishes they wanted us to have.

You may be thinking that, to some extent, we do live in such a world, where we are free to make numerous choices, but are manipulated into making many of our choices by advertising, television, public relations, spin doctors, salespersons, marketers, and sometimes even by friends, parents, relatives, rivals, or enemies. One indication of how important free will is to us is that people generally feel revulsion at such manipulation. When people find out that what they thought were their own wishes were actually manipulated by others who wanted them to choose in just the way they did, they feel demeaned. Such situations are demeaning because we realize we were not our own persons; and having free will is about being your own person.

The problem is brought out in a striking way by twentieth-century utopian novels, such as Aldous Huxley's *Brave New World* and B. F. Skinner's *Walden Two.*[2] In the fictional societies described in these famous works, people can have and do what they will or choose, but only to the extent that they have been conditioned by behavioral engineers or neuro-chemists to will or choose what they can have and do. In *Brave New World,* the lower-echelon workers are under the influence of powerful drugs so that they do not dream of things they cannot have. They are quite content to play miniature golf all weekend. They can do what they want, though their wants are meager and controlled by drugs.

The citizens of Skinner's *Walden Two* have a richer existence than the workers of *Brave New World.* Yet their desires and purposes are also covertly controlled, in this case by behavioral engineers. Walden Two-ers live collectively in a kind of rural commune; and because they share duties of farming and raising children, they have plenty of leisure. They pursue arts, sciences, crafts, engage in musical performances, and enjoy what appears to be a pleasant existence. The fictional founder of Walden Two, a fellow named Frazier, forthrightly says that their pleasant existence is brought about by the fact that, in his community,

persons can do whatever they want or choose because they have been behaviorally conditioned since childhood to want and choose only what they can have and do. In other words, they have maximal *surface freedom* of action and choice (they can choose or do anything they want), but they lack a *deeper freedom* of the will because their desires and purposes are created by their behavioral conditioners or controllers. Their wills are not of "their own" making. Indeed, what happens in Walden Two is that their surface freedom to act and choose as they will is maximized by minimizing the deeper freedom to have the ultimate say about what they will.

Thus Frazier can say that Walden Two "is the freest place on earth" (p. 297), because he has surface freedoms in mind. For there is no *coercion* in Walden Two and no *punishment* because no one has to be forced to do anything against his or her will. The citizens can have anything they want because they have been conditioned not to want anything they cannot have. As for the deeper freedom, or free will, it does not exist in Walden Two, as Frazier himself admits (p. 257). But this is no loss, according to Frazier. Echoing *Walden Two*'s author, B. F. Skinner (a foremost defender of behaviorism in psychology), Frazier thinks the deeper freedom of the will is an illusion in the first place. We do not have it anyway, inside or outside Walden Two. In our ordinary lives, he argues, we are just as much the products of upbringing and social conditioning as the citizens of Walden Two, though we may delude ourselves into thinking otherwise. The difference is that, unlike Walden Two, our everyday conditioning is often haphazard, incompetent, and harmful.

Why then, Skinner asks, reject the maximal surface freedom and happiness of Walden Two for a deeper freedom of the will that is something we do not and cannot have anyway? Along with many other scientists, he thinks the idea that we could be *ultimate* determiners of our own ends or purposes (which is what the deeper freedom of the will would require) is an impossible ideal that cannot fit into the modern scientific picture of the world. To have such freedom, we would have to have been the original creators of our own wills—causes of ourselves. But if we trace the psychological springs of action back further and further to childhood, we find that we were less free back then, not more, and more subject to conditioning. We thus delude ourselves into thinking that we have sacrificed some real (deeper) freedom for the happiness of Walden Two. Rather we have gained a maximum amount of the only kind of freedom we really can have (surface freedom), while giving up an illusion (free will).

Seductive as these arguments may be, there are many people (myself included) who continue to believe that something important is missing in Walden Two and that the deeper freedom is not a mere illusion. Such persons want to be the ultimate designers of their own lives as Frazier was for the lives of Walden Two. They want to be the creators, as he was, not the pawns—at least for their own lives. What they long for is what was traditionally meant by "free will."

Here is yet another way of looking at it. Free will in this deeper sense is also intimately related to notions of moral responsibility, blameworthiness, and praiseworthiness. 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 on a daily basis. At first, our thoughts of the young man are filled with anger and resentment. But as we listen daily to how he came to have such a mean character and perverse motives—a sordid 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 influenced him. We begin to feel angry with them as well as him. 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 about what may be called *ultimate responsibility.* We know that parenting and society, genetic makeup 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's what we wanted to know about the young man. The question of whether he is merely a victim of his bad circumstances or has some residual responsibility for being what he is depends on whether these orher factors were or were not *entirely* determining.[3]

Turning this around, if there were factors or circumstances that entirely determined what he did, then to be ultimately responsible, he would have had to be responsible to some degree for some of those factors by virtue of earlier acts through which he formed his present character. As the philosopher Aristotle put it centuries ago, if a man is responsible for the wicked acts that flow from his character, then he must at one time in the past have been responsible for forming the character from which these acts flow. But, of course, if *all* of our choices and actions were entirely determined by prior circumstances, we would have had to be responsible to some degree for some of these earlier circumstances by still earlier acts of ours, and so on indefinitely backward in time—an impossibility for finite creatures like ourselves. At some point, if we are to be ultimately responsible for being what we are, there must be acts in our life histories in which parenting and society, genetic make-up, and other factors did not completely determine how we acted, but left something over for us to be responsible for then and there. This is why many people have thought that the deeper freedom of the will is not compatible with being completely determined by the past. Surface freedoms (to do or choose what we will) may be compatible with determinism, but free will does not seem to be (as Skinner himself realized).

II

Yet such thoughts only lead to a further problem that has haunted free will debates for centuries: If this deeper freedom of the will is not compatible with determinism, it does not seem to be compatible with *indeterminism* either. An event that is undetermined might occur or might not occur, given the entire past. (A determined event *must* occur, given the entire past.) Thus, whether or not an undetermined event actually occurs, given its past, is a matter of chance. But chance events occur spontaneously and are not under the control of anything, hence not under the control of agents. How then could they be free and responsible actions? If, for example, a choice occurred by virtue of a quantum jump or other undetermined event in your brain, it would seem a fluke or accident rather than a responsible choice. Undetermined events in the brain or body, it seems, would inhibit or interfere with freedom, occurring spontaneously and not under our control. They would turn out to be a nuisance—or perhaps a curse, like epilepsy—rather than an enhancement of our freedom.

Or look at the problem in another way that goes a little deeper. If my 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. This is what indeterminism and the denial of determinism mean: exactly the same past, different outcomes. Imagine, for example, that I had been deliberating about where to spend my vacation, in Hawaii or Colorado, and after much thought and deliberation had decided I 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 to my favoring and choosing Hawaii over Colorado, might by chance have resulted in my 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. Because I had come to favor Hawaii and was about to choose it, when by chance I chose Colorado, I would wonder what went wrong in my brain and perhaps consult a neurologist.

For reasons such as these, people have argued that undetermined free choices would be "arbitrary," "capricious," "random," "irrational," "uncontrolled," "inexplicable," or merely "matters of luck or chance," not really free and responsible choices at all. If free will is not compatible with determinism, it does not seem to be compatible with indeterminism either.

These charges are powerful ones and defenders of free will over the centuries have made extraordinary claims attempting to respond to them. Free will does require indeterminism, these defenders have said. But it cannot *merely* be indeterminism or chance. Some "extra factors" must be involved in free will that go beyond ordinary scientific or causal understanding. Immanuel Kant, for example, insisted that we can't explain free will in scientific and psychological terms. To account for it we have to appeal to the agency of what he called a "noumenal self" outside space and time that could not be studied in scientific terms.[4] Others have appealed to what Nobel physiologist John Eccles calls a "transempirical power center," which would intervene in the brain, filling the causal gaps left by indeterminism or chance.[5] Still others have appealed to a special kind of agent-causation—or, as Roderick Chisholm has called it, "immanent causation"—that cannot be explained in terms of the ordinary scientific modes of causation in terms of events or occurrences. Where all prior events, both physical and mental, leave a choice or action undetermined, the agent- or immanent cause determines it, but cannot be determined in turn because it is not an event. The agent-cause is, in Chisholm's words, a "prime mover unmoved."[6]

Such unusual strategems are common among defenders of an indeterminist free will (who often nowadays are called "incompatibilists" because they believe that free will is not compatible with determinism and "libertarians" because they believe in addition that free will is not an illusion). But these unusual strategems, such as noumenal selves, transempirical power centers, and agent- or immanent causes, have unfortunately reinforced the view, now widespread among philosophers and scientists, that traditional notions of free will requiring indeterminism are mysterious and have no place in the modern scientific picture of the world. Such libertarian strategems, to their critics, are reminiscent of the old debates about vital forces in the biology of the nineteenth century, where obscure forces were postulated to explain what otherwise could not be explained about living things. They remind us of the Arkansas farmer when he first saw an automobile. He listened intently to the explanation of how the internal combustion engine worked, and nodded in agreement, but insisted on looking under the hood anyway because, as he said, "there must be a horse in there somewhere."

Thus, defenders of a nondeterminist free will are faced with a dilemma that was expressed by philosopher Thomas Hobbes at the beginning of the modern era. When trying to explain free will, these incompatibilist or libertarian defenders tend to fall either into "confusion" or "emptiness"—the confusion of identifying free will with indeterminism or the emptiness of mysterious accounts of agency in terms of noumenal selves, transempirical power centers, non-occurrent or agent-causes, or other strategems whose operations remain obscure and unexplained. What is needed to escape this dilemma is some new thinking about how free will can be reconciled with indeterminism and how it might fit into the modern scientific picture of the world, without appealing to extra factors that have made it seem so mysterious. In the remainder of this essay, I want to suggest some new ways of thinking about this problem and about free will generally, which may stir you to do likewise.[7]

III

The first thing to note is that indeterminism does not have to be a factor in all acts done "of our own free wills." Not all of them have to be undetermined. Frequently in everyday life we act from existing motives without having to think or deliberate about what to do. At such times, we may very well be determined by our existing characters and motives. Yet we may also at such times be acting "of our own free wills" to the extent that we formed our present characters and motives (our own wills) by earlier choices or actions that were not themselves determined. Recall again Aristotle's claim that if a man is responsible for the wicked acts that flow from his character, he must at one time in the past have been responsible for forming the character from which these acts flow. Not all choices or acts done "of our own free wills" have to be undetermined, but only those choices or acts in our lifetimes by which we made ourselves

into the kinds of persons we are. Let us call these "self-forming choices or actions" or SFAs.

I believe that such undetermined self-forming choices and actions (SFAs) occur at those difficult times of life when we are torn between competing visions of what we should do or become, and that they are more frequent than we think. 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. At such times, there is tension and uncertainty in our minds about what to do. I suggest that this 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 would thus be reflected in the indeterminacy of our neural processes themselves. What is experienced personally as uncertainty corresponds physically to the opening of a window of opportunity that temporarily screens off complete determination by influences of the past. (By contrast, when we act from predominant motives or settled dispositions, the uncertainty or indeterminacy is muted. If it did play a role in such cases, it would be a mere nuisance or fluke, as critics suggest, like the choice of Colorado when we favored Hawaii.)

When we do decide under such conditions of uncertainty, the outcome is not determined because of the preceding indeterminacy—and yet it 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 a conflict of this kind. She is on the way to a business meeting important to her career when she observes an assault taking place in an alley. An inner struggle ensues between her moral conscience telling her to stop and call for help, and her career ambitions telling her she cannot miss this meeting. She has to make an effort of will to overcome the temptation to go on to her meeting. 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 because, while she wanted to overcome temptation, she also wanted to fail, for quite different and incommensurable reasons. When we, like the businesswoman, 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 does not necessarily undermine rationality and voluntariness, so indeterminism, in and of itself, does not necessarily 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 with background noise or distraction. Whether you are going to succeed in solving the mathematical problem is uncertain and undetermined because of the distracting indeterministic 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 distracting neural noise would have been an obstacle that you overcame by your effort.

There are numerous other examples supporting this point, where indeterminism functions as an obstacle to success without precluding responsibility. Consider an 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 obviously yes because he intentionally and voluntarily succeeded in doing what he was *trying* to do—kill the prime minister. Yet his action, killing

Difficult choice situation; conflict of will

Solving math problem example
Assassin / Irate husband:
Not SFAs but cases of indeterministic responsibility

the prime minister, was undetermined. One might even say "he got lucky" in killing the prime minister, because there was a chance he might have missed. Yet, for all that, he *did* kill the prime minister and *was* responsible for it.

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, intending to break it. Again, we suppose that some indeterminism in the nerves of his arm makes the momentum of his swing indeterminate so that it is literally not determined whether the table will break right up to the moment when it is struck. Whether the husband breaks the table or not is undetermined and yet he is clearly responsible if he does break it. (It would be a poor excuse for him to say to his wife "chance did it, not me" or "it wasn't my doing; it happened by chance." She would not be impressed.)

To be sure, such examples—of the mathematical problem, the assassin, and the husband—do not amount to genuine exercises of free will in "self-forming actions" or SFAs, such as the businesswoman's, where the wills of the agents are divided between conflicting motives. The businesswoman wants to do the right thing and help the victim, but she also wants to go on to her meeting. By contrast, the will of the assassin is not equally divided. He wants to kill the prime minister, but does not also want to fail. (If his conscience bothered him and he was undecided about what to do up to the last minute, that would be another matter. *Then* his choice would be a self-forming action or SFA, like the businesswoman's. But such was not the case.) Thus, if the assassin fails to hit his target, it will be merely by chance or as a fluke, not voluntarily (and so also for the husband and mathematical problem-solver). Cases such as the assassin, husband, and mathematical problem-solver are therefore not all that we want. Yet they are a step in the right direction because they show that indeterminism does not necessarily rule out action and responsibility, any more than it necessarily rules out rationality and voluntariness. To go further, we have to dig more deeply and add some further ideas.

IV

Let us imagine in cases of self-forming choices, like the businesswoman's, where there *is* conflict in the will, that the indeterministic noise that is providing an obstacle to her overcoming temptation (and stopping to help the victim) is not coming from an external source, but is coming from her own will, because she also deeply desires to do the opposite (go on to her meeting). Imagine that in such conflicting circumstances, two competing (recurrent) neural networks are involved. (These are complex networks of interconnected neurons in the brain circulating impulses in feedback loops that are generally involved in high-level human cognitive processing.[8]) The input of one of these networks is coming from the woman's desires and motives for stopping to help the victim. If the network reaches a certain activation threshold (the simultaneous firing of a complex set of "output" neurons), that would represent her choice to help. For the competing network, the inputs are her ambitious motives for going on to her meeting, and its reaching an activation threshold would represent the choice to go on. (If one network activates, the other will be inhibited and the contrary choice will not be made.)

Now imagine further that these two competing networks are connected so that the indeterministic noise that is an obstacle to her making one of the choices is coming from her desire to make the other. Thus, as suggested for self-forming choices or SFAs, the indeterminism arises from a *tension-creating conflict in the will*. In such circumstances, when either of the pathways "wins" (i.e., reaches an activation threshold, which amounts to choice), it will be like the agent's solving the mathematical problem by overcoming the indeterministic background noise generated by the other. And just as we could say, when you solved the mathematical problem by overcoming the distracting noise through your effort, that you did it and are responsible for it, so one can say this as well, I would argue, in the present case, *whichever one is chosen*. The neural pathway through which she succeeds in reaching a choice threshold will have overcome

16

the obstacle in the form of indeterministic noise coming from the other pathway.

Note that, in these circumstances, the choices either way will not be "inadvertent," "accidental," "capricious," or "merely random," because they will be *willed* by the woman either way, when they are made, and done for *reasons* either way (moral convictions if she turns back, ambitious motives if she goes on) which she then and there endorses. And these are the conditions usually required to say something is done "on purpose," rather than accidentally, capriciously, or merely by chance. Moreover, these conditions taken together (that she wills it, and does it for reasons, and could have done otherwise willingly and for reasons) rule out each of the normal motives we have for saying that agents act, but do not have control over their actions (coercion, constraint, inadvertence, mistake, and control by others). None of these obtain in the businesswoman's case. She is not coerced (no one is holding a gun to her head), not physically constrained or disabled, not forced or controlled by others; nor does she act inadvertently or by mistake, but on purpose either way, as just noted.

Of course, with "self-forming" choices of these kinds, agents cannot control or determine which choice outcome will occur *before* it occurs or the outcomes would be *pre*determined after all. (That would be like deciding beforehand what you are going to decide.) But it does not follow that, because one does not control or determine which of a set of outcomes is going to occur before it occurs, one does not control which of them occurs, *when* it occurs. When the above conditions for self-forming choices are satisfied, agents exercise control over their future lives *then and there* by deciding. Indeed, they have what may be called "plural voluntary control" in the following sense: Agents have plural voluntary control over a set of options (stopping to help or going on to a 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 by mistake or accident, without being coerced or compelled in doing so, or otherwise controlled by other agents or mechanisms. We have seen that each of these conditions can be satisfied in cases of SFAs, like the businesswoman's, despite the indeterminism involved.[9] These conditions of plural voluntary control may be summed by saying, as people often do, that the agents can choose either way "at will." ("Plural" in "plural voluntary control" means "more-than-one-way" and "voluntary" means "in accordance with one's will.")

Note also that this account of self-forming choices amounts to a kind of "doubling" of the mathematical problem. It is as if an agent faced with such a 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 coming from the other, 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, they will have succeeded in doing what they were knowingly and willingly trying to do. Recall again the cases of 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, 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 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 a complex creature, 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 as *her* resolution of the conflict in her will, voluntarily and intentionally, not by accident or mistake.

V

Yet it is still hard to shake the intuition that if choices are undetermined, they *must* happen merely by chance—and so must be "random," "capricious," "uncontrolled," "irrational," "inexplicable," and all the other things charged. I do not deny the powerful hold such intuitions have upon us. They are among the reasons why free will continues to be such a deep problem, even for those who want to believe in it. But the very fact that it has been such a problem for so long should also suggest that we cannot take ordinary intuitions about free will at face value without questioning them. If we are ever going to understand it, we will likely have to break old habits of thought and learn to think in new ways.

The first step in doing this is to question the intuitive connection in most 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 connotation of "it's out of my control." So using them already begs certain questions, whereas "indeterminism" is a technical term that merely precludes *deterministic* causation, though not causation altogether. Indeterminism is consistent with non-deterministic or probabilistic causation, where the outcome is not inevitable. It is therefore a mistake (alas, one of the most common in debates about free will) to assume that "undetermined" means "uncaused."

Indeterministic causation (= agent causation?)

Another source of misunderstanding is this: Because 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 temptation (to go on to her meeting) and then at the last instant "chance taking over" and deciding the issue for her. But this image is misleading. On the view just described, 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.)

And just as expressions such as "she chose *by* chance" can mislead us in such contexts, so can expressions like "she got lucky." Recall that in the cases of the assassin and the husband, one might say "they got lucky" in killing the prime minister and breaking the table because their actions were undetermined. Yet, as we noted, it does not follow that they were not responsible. So ask yourself this question: Why does the inference "he got lucky, *so he was not responsible?*" fail when it does fail, in the cases of the husband and the assassin? The first part of an answer has to do with the point made earlier that "luck," like "chance," has 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

17

assassin and husband cases, which *is* implied by indeterminism, I suggest, 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 further meanings in the husband and assassin cases that are often associated with "luck" and "chance" in ordinary usage (for example, the outcome was not his doing, or occurred by *mere* chance, or he was not responsible for it), the inference would not fail for the husband and assassin, as it clearly does. 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" fails for 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, that is to say, 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), first, she will have "succeeded *despite the probability or chance of failure;* second, 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 third, 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 it 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.

Let us try another tack. Perhaps we are begging the question by assuming at the outset that the outcomes of the woman's efforts are her *choices*. If they are not choices to begin with, they cannot be voluntary choices. One might argue this on the grounds that (A) "if an event is undetermined, it must be something that merely *happens* and cannot be somebody's choice"; or (B) "if an event is undetermined, it must be something that merely happens, it cannot be something an agent *does* (it cannot be an action)." But to see how question-begging these assumptions are, one has only to note that A and B imply respectively (A′) "if an event is a choice, it must be determined" ("all choices are determined") and (B′) "if an event is an action, it must be determined" ("all actions are determined"). Are these claims supposed to be true necessarily or by definition? If so, the free will issue would be solved by fiat; it would follow merely from the meanings of the words that all choices and actions are determined.

But why should we believe this? Was the husband's breaking the table not something he did because it was not determined? Recall that "undetermined" does not mean "uncaused." The breaking of the table was caused by the swing of his arm, and though the outcome was not inevitable, that was good enough for saying he did it and was responsible. As for choices, a choice is the formation of an intention or a purpose to do something. It resolves uncertainty and indecision in the mind about what to do, "setting the mind" on one alternative rather than another. Nothing in such a description implies that there could not be some indeterminism in the deliberation and neural processes of an agent's preceding choice corresponding to the agent's uncertainty about what to do. Recall from preceding arguments that the presence of indeterminism does not mean the outcome happened merely by chance and not by the agent's effort.

But it is one thing to choose, in the sense of forming an intention; it is another thing to have *control* over one's choosing. Perhaps this is where the real problem lies. Would not the presence of indeterminism at least *diminish* the control persons have over their choices and other actions? 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? Moreover, this limitation is connected with another often noted by critics—that indeterminism, wherever it occurs, seems to be a *hindrance* or *obstacle* to our realizing our purposes and hence an obstacle to our freedom.

These concerns are closer to the mark, and there is something to them. But rather than being devastating objections to an incompatibilist account of free will, I think they reveal something important about such a free will. I think 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 businesswoman (and for 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).

So, 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. As a consequence, whichever choice she makes, whichever effort wins out, she will have to overcome the hindrance or obstacle provided by the indeterminism coming from the other. 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 that would stand in the way of her choosing it. 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. We can concede then that indeterminism is a hindrance and a nuisance, but a necessary one if we are to have ultimate responsibility for our own wills. Being "your own self" is a struggle. We can appreciate why existentialist philosopher Jean-Paul Sartre said that true freedom (free will) is a burden many people want to "escape"—preferring instead that others tell them what to do and how to live, or perhaps preferring that their choices always be easy.[10] In an earlier time, St. Augustine asked why God would have given us free will, since it is such a pain to us and to others; and the answer was that without it we would lack the greater good of being ultimately responsible for what we are and what we do.[11]

Perhaps we should look in another direction that has also led to doubts about whether free will can be reconciled with indeterminism or chance. What might be going in the brain, we might ask, when free choices take place? If neuroscientists were to inspect the woman's brain when she was struggling with her moral decision, wouldn't it be the case that they would find nothing more than interconnected sets of neuron firings in which micro-indeterminacies were not negligible? These interconnected neuron firings would in turn terminate in some definite configuration of nerve firings that corresponded to the "choice" to stop and help the victim or in another set of firings corresponding to the "choice" to go on to her meeting. But why one of these outcomes occurred rather than the other would be inexplicable in terms of the preceding processes. Probabilities could be assigned for one outcome rather than the other, but that is all. And this looks like chance.

I agree that if the physical descriptions of these events were the only legitimate ones, then free will would look like nothing more than chance or

probability. When neuroscientists described it in physico-chemical terms, all they would get are indeterministic chaotic processes with probabilistic outcomes. In short, if described from a physical perspective alone, free will looks like chance. But the physical description is not the only one to be considered. The indeterministic chaotic process is also, experientially considered, the agent's effort of will—something the agent is doing. And the undetermined outcome of the process, one way or the other, is experientially the agent's choice—something the agent does, not something that merely happens to the agent. So viewed from another perspective, the neural output that represents the choice is the result of the agent's effort even though the outcome is not determined.

If we did not add these mental descriptions of what is going on to the physical descriptions, something important would be left out of our picture of the world. To make sense of free will, we do not have to be complete (substance) dualists about mind and body, as Descartes was. But we cannot be extreme "eliminative" materialists either. We can't expect to lop off from the top of our world-view all psychological descriptions of human beings in terms of beliefs, desires, intentions, efforts, choices, and consciousness (leaving only descriptions in neurophysiological terms), *and expect free will to survive*. The fact is that a lot of other things important to us would not survive either if we were to lop off these psychological descriptions from our descriptions of the world, such as personhood, rationality, subjectivity, morality, and so on.

But notice that *this* problem is not a special one for theories of free will that presuppose indeterminism. Suppose you believed that all choices and actions were determined and that human free agency was compatible with determinism. You still could not adequately describe human agency, if you confined yourself to describing the brain in chemical and neurophysiological terms alone, leaving out all ordinary psychological descriptions in terms of beliefs, desires, intentions, efforts, choices, and consciousness. Determinists and compatibilists about free will cannot eliminate supervenient mental or psychological descriptions either, if they are going to describe human agency. It is no less a mystery how neural firings in the brain could be, or give rise to, conscious beliefs, efforts, or choices if these neural firings are *determined* than if they are undetermined. This problem (the problem of consciousness, or more generally, the "mind/body problem") is no special problem for indeterminist theories of free will like the one given here. It is a problem for anyone who wishes to talk about free agency, whatever position they take on free will, compatibilist or incompatibilist, determinist or indeterminist.

These reflections naturally raise the further question of whether the indeterminism that is required by an incompatibilist theory of free will is actually there in the brain. This is an empirical question that can only be decided by scientific research and not by a philosophical theory or armchair speculation (much as philosophers would like to decide all questions *a priori*, or before all experience). While we cannot resolve this question, we can at least keep our minds open about it. There is so much more to be learned about the brain and living things. One caution, however: If you are inclined to believe that free will is incompatible with determinism (if you are an incompatibilist or libertarian), don't think you can escape such scientific and empirical questions altogether unless you want to leave free will a complete mystery. Even if you appealed to "transempirical power centers" or "non-event" agent causes to make sense of free will (as libertarians often do), there would still have to be some indeterminacy in the natural world—and presumably in the brain where it counts—to make room in nature (to provide the "causal gaps") for the intervention of these additional causes or agencies. As the ancient Epicurean philosophers said centuries ago, if the atoms do not sometimes "swerve" in undetermined ways, there will be no room in nature for free will.

Addressing this problem earlier, I suggested that conflicts in the wills of agents associated with self-forming choices might "stir up chaos" in the brain, sensitizing it to quantum indeterminacies at the neuronal level, which would then be magnified to affect the neural networks as a whole. This is speculative to be sure, and others writers have suggested different ways in which indeterminacy might be involved in the brain.[12] But such speculations are not merely idle. There is some evidence that unpredictable chaotic activity plays a role in the brain and human cognition, providing some of the flexibility that the nervous system needs to react creatively to an ever-changing environment. A recent article in the journal *Behavior and Brain Sciences*, entitled "How Brains Make Chaos in Order to Make Sense of the World", defends this role, as do other recent writings.[13] Now it is true that chaos (or chaotic behavior) in physical systems, though unpredictable, is nonetheless usually deterministic. Chaos does not of itself imply indeterminism. But chaotic behavior in physical systems does involve "sensitivity to initial conditions." Minute differences in the initial conditions of chaotic physical systems, including living things, may be magnified, giving rise to large-scale, undetermined effects. If the brain does "make chaos to understand the world," its sensitivity to initial conditions may magnify quantum indeterminacies in neural networks whose outputs can depend on minute differences in the timings of individual neuron firings. So while quantum physics and the new sciences of chaos and complexity may not give us the indeterminism needed for free will alone (because the uncertainty of the former is usually negligible in larger physical systems and the latter need not be indeterministic by itself), they might do so together.

In any case, I have not tried to settle such empirical questions, nor could I. What I have been addressing is another set of questions that incline people to write off incompatibilist views of free will from the start, believing they could not possibly make sense and could not be reconciled with the modern scientific picture of human beings, *even if* indeterminism were somehow available in the physical world. I have argued to the contrary that if the indeterminism is there in nature, then something could be done to make sense of free will.

Let me conclude with one final objection that 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 because the agents cannot in principle have sufficient or overriding *prior* reasons for making one option and one set of reasons prevail over the other. The agents *make* one set of reasons prevail *by* choosing, to be sure, but they could as well have made the other set of reasons prevail by choosing differently.

I agree that there is some truth to this charge as well. But I would argue that such arbitrariness relative to prior reasons also 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 is consistent with my past and is one branching pathway 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*), they are "arbiters" of their own lives, "making themselves" out of a past that, if they are truly free, does not limit their future pathways to one.

Suppose we were to say to them, "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 they 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 described above, I am in the process of writing an unfinished story and forming an unfinished character who, in my case, is myself."

NOTES

1. *Paradise Lost* (London: Methuen, 1955), prologue.

2. *Brave New World* (San Francisco: HarperCollins, 1989). *Walden Two* (New York: MacMillan, 1962). Page references in the paper to *Walden Two* are to this edition.

3. This is why we are naturally inclined to ask in cases like this whether someone else in exactly the same circumstances might have acted differently.

4. Kant, *The Critique of Practical Reason*. Trans. by L. W. Beck (Indianapolis: Bobbs-Merrill, 1956), part III.

5. Eccles, *Facing Reality* (New York: Springer-Verlag, 1970).

6. For defenses of this agent-causal position by various authors, see the essays in T. O'Connor (ed.), *Agents, Causes and Events: Essays on Indeterminism and Free Will* (Oxford: Oxford University Press, 1995).

7. These ideas are developed at greater length in my book, *The Significance of Free Will* (Oxford and New York: Oxford University Press, 1996; paperback edition, 1998).

8. A readable and accessible introduction to the role of neural networks (including recurrent networks) in cognitive processing is P. M. Churchland, *The Engine of Reason, the Seat of the Soul* (Cambridge, MA: MIT Press, 1996).

9. I show in greater detail that each of these conditions can be satisfied by self-forming choices or SFAs in *The Significance of Free Will* (op. cit.), chapter 8.

10. Sartre, "Selections from *Being and Nothingness*." In S. Morgenbesser et al. (eds.), *Free Will* (Englewood Cliffs, NJ: Prentice-Hall, 1962), pp. 95–113.

11. Augustine, *On the Free Choice of the Will* (Indianapolis: Bobbs-Merrill, 1964), Part I.

12. For example, H. Stapp, *Mind, Matter and Quantum Mechanics* (New York: Springer-Verlag, 1993); D. Hodgson, *The Mind Matters* (Oxford: Clarendon Press, 1991); R. Penrose, *The Emperor's New Mind* (Oxford: Oxford University Press, 1989); J. Eccles (op. cit.)

13. The article is C. Skarda and W. Freeman, *Behavioral and Brain Sciences* 10 (1987): 161–195; other writings on the subject include A. Babloyantz and A. Destexhe, "Strange Attractors in the Human Cortex." In L. Rensing (ed.), *Temporal Disorder in Human Oscillatory Systems* (New York: Springer-Verlag, 1985); G. Scott and M. McMillen (eds.), *Dissipative Structures and Spatiotemporal Organization Studies in Biomedical Research* (Ames: Iowa State University Press, 1980); H. Walter, *Neurophilosophy and Free Will*. Trans. by Cynthia Stohr (Cambridge MA: MIT Press, 2001), Part III.

Freedom and Moral Responsibility

The Impossibility of Moral Responsibility*

GALEN STRAWSON

Galen Strawson teaches philosophy at the University of Oxford.

I

THERE IS AN ARGUMENT, which I will call the Basic Argument, which appears to prove that we cannot be truly or ultimately morally responsible for our actions. According to the Basic Argument, it makes no difference whether determinism is true or false. We cannot be truly or ultimately morally responsible for our actions in either case.

The Basic Argument has various expressions in the literature of free will, and its central idea can be quickly conveyed. (1) Nothing can be *causa sui*—nothing can be the cause of itself. (2) In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects. (3) Therefore nothing can be truly morally responsible.

In this paper I want to reconsider the Basic Argument, in the hope that anyone who thinks that we can be truly or ultimately morally responsible for our actions will be prepared to say exactly what is wrong with it. I think that the point that it has to make is obvious, and that it has been underrated in recent discussion of free will—perhaps because it admits of no answer. I suspect that it is obvious in such a way that insisting on it too much is likely to make it seem less obvious than it is, given the innate contrasuggestibility of human beings in general and philosophers in particular. But I am not worried about making it seem less obvious than it is so long as it gets adequate attention. As far as its validity is concerned, it can look after itself.

A more cumbersome statement of the Basic Argument goes as follows.

(1) Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to 'reflex' actions or mindlessly habitual actions).

(2) When one acts for a reason, what one does is a function of how one is, mentally speaking. (It is also a function of one's height, one's strength, one's place and time, and so on. But the mental factors are crucial when moral responsibility is in question.)

(3) So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking—at least in certain respects.

(4) But to be truly responsible for how one is, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way.

*From Philosophical Studies 75: 5–24.

20