Philos Stud (2009) 144:35–44
DOI 10.1007/s11098-009-9365-y

Libertarianism

Robert Kane

Published online: 27 March 2009

My purpose in this symposium, part of a discussion of four competing views of free will is to defend the libertarian view of free will that I have been developing over the past several decades. According to that view, free will is incompatible with determinism. I argue for this incompatibility on grounds that free will requires ultimate responsibility for our actions, i.e., the power to be ultimate originators of at least some of our own ends or purposes. I argue further that this view of free will can be reconciled with modern science against criticisms that it does not allow sufficient control, reduces to mere chance or luck, and other criticisms. In the process, I discuss questions about the nature of action, moral responsibility, choice, effort, rationality and other notions commonly associated with the freedom of the will, free will, moral responsibility, incompatibilism, libertarian view of free will, effort, will, self-forming actions (SFAs), alternative possibilities, luck, chance, Frankfurt-type examples.

To make sense of libertarian free will, as I argue in the Four Views volume, and have argued elsewhere, you must learn to think in new ways, to think outside the box, so to speak. Otherwise such a free will will appear utterly mysterious, the greatest self-contradiction conceived by the mind of man, as Nietzsche inimitably put it; and the objections forcefully raised against libertarian views in the volume by Fischer, Pereboom and Vargas and by them and by McKenna in the papers that follow will appear overwhelming. So let's see how this outside the box thinking would go.

First one must note that the Compatibility Question as usually framed—"Is freedom compatible with determinism?"—is altogether too simple and ill-formed. The reason is that there are many meanings of "freedom" and many of them *are* compatible with determinism (as McKenna suggests). Libertarians should simply

R. Kane (✉)
Department of Philosophy, The University of Texas at Austin, Waggener Hall 316, Austin, TX 78712-1180, USA
e-mail: rkane@uts.cc.utexas.edu

concede this point and move on. What libertarians should insist upon is that there is *at least one* kind of freedom worth wanting that is incompatible with determinism. This significant further freedom is "free *will*," which I define as "the power to be the ultimate creator and sustainer of one's own ends or purposes." To say this further freedom is important is not to deny the importance of everyday compatibilist freedoms from coercion, compulsion, political oppression, and the like; it is only to say that human longings transcend them.

Consider, second, that debates about incompatibilism have tended to focus on the requirement that free agents "could have done otherwise," the requirement I call AP for alternative possibilities. Debates about whether free will and moral responsibility require alternative possibilities and whether alternative possibilities are compatible with determinism have tended to stalemate, as we know, over differing interpretations of "can," "power," "ability" and "could have done otherwise." And I believe that there were good reasons for these stalemates having to do with the different meanings of freedom just mentioned. That is why I argue that to resolve the compatibility issue, one must look beyond alternative possibilities. (On this point, I agree with Fischer, Pereboom and McKenna, though we disagree on details, as we'll see.)

One must look to another condition that I call ultimate responsibility or UR, the basic idea of which is this: to be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient reason, cause or motive for the action's occurring. If, for example, a choice issues from, and can be sufficiently explained by, an agent's character and motives (together with background conditions), then to be *ultimately* responsible for the choice, the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past for having the character and motives he or she now has. Compare Aristotle's claim that if a man is responsible for wicked acts that flow from his character, he must at some time in the past have been responsible for forming the wicked character from which these acts flow.

This UR condition accounts for the "ultimate" in the definition of free will: "the power of agents to be the *ultimate* creators and sustainers of their own ends or purposes." Now UR does not require that we could have done otherwise (AP) for *every* act done of our own free wills—thus vindicating philosophers such as Frankfurt, Fischer, Pereboom, McKenna and others who insist that we can be held morally responsible for many acts even when we could not have done otherwise. But the vindication is only partial. For UR *does* require that we could have done otherwise with respect to *some* acts in our past life histories by which we formed our present characters. I call these "self-forming actions," or SFAs. Often in everyday life, we act "of our own free will" in the sense of a will already formed. But on such occasions, the will (i.e., character, motives and purposes) from which we act is "*our own* free will" to the extent that we had a role in forming it by earlier SFAs that were not determined and with respect to which we could have voluntarily and rationally done otherwise. If this were not so, there would have been *nothing we could have ever done differently in our entire lifetimes to make our wills other than they are*; and we would never act "of our own free will" in the sense of a will ultimately of our own making.

These reflections suggest a third point that has been lost sight of in the modern era: the free will issue really is about the freedom *of the will* and not merely about freedom of action. To reduce the problem to one merely of freedom of action is to oversimplify it. Free will is not just about free action. It is about *self-formation*, about the formation of our "wills" or how we got to be the kinds of persons we are, with the characters, motives and purposes we now have. Were *we* ultimately responsible to some degree for having the wills we do have, or can the sources of our wills be completely traced backwards to something over which we had no control, such as Fate or the decrees of God, or heredity and environment or social conditioning or hidden controllers, and so on? Therein, I believe, lies the core of the traditional problem of "free will."

If the case for incompatibility cannot be made on AP alone, it can be made if UR is added. If agents must be responsible to some degree for anything that is a sufficient cause or motive for their actions, an impossible infinite regress of past actions would be required unless *some* actions in the agent's life history (self-forming SFAs) did not have either sufficient causes or motives (and hence were *undetermined*).

We may sum this up in terms of three kinds of free acts. Free acts may be

(1) acts done voluntarily, on purpose and for reasons that are not coerced, compelled or otherwise constrained or subject to control by other agents.
(2) acts [which are free in sense 1 that are also] done "of our own free will" in the sense of a will that we are ultimately responsible (UR) for forming.
(3) "self-forming" acts (SFAs) or "will-setting" acts by which we form the will from which we act in sense 2.

Acts of type 1 are compatible with determinism. One can act freely, in the sense of voluntarily, on purpose and for reasons, without being coerced, compelled or otherwise constrained or controlled by others, even if determinism should be true and even if one's act is determined. Type 1 freedom is thus a compatibilist freedom; and as I said earlier there are important kinds of freedom that *are* compatible with determinism.

Free acts of types 2 and 3 by contrast are incompatibilist or libertarian free acts. They could not exist in a determined world. But only acts of type 3 have to be undetermined. Acts done of our own free will of type 2 may be determined (though they need not be) and may even be such that the agents could not have done otherwise. Yet they are libertarian free because even though type 2 acts of free will may themselves be determined, they could not exist *in a determined world* because they presuppose other acts (of type 3) that are not determined.

Type 3 acts are the "self-forming" or "will-setting" acts—SFAs as I call them—by which we form in an undetermined way the wills from which we later act. Consequently, I view freedoms of type 2 and 3 as defining *free will*, while freedom of type 1 defines freedom of *action*. Moreover, the three types of acts are nested. Type 3 acts ("self-forming acts" of free will) are also free in sense 2 (they are ultimately responsible acts of free will, albeit of a special kind). And acts of types 2 *and* 3 (acts done of our own free will) are also free acts of type 1 (they must be voluntary, uncoerced, non-compelled, etc.). So *freedom of will* (of types 2 and 3) *is*

a kind of freedom of action (of type 1), albeit a special kind; and incompatibilist freedom (of types 2 and 3) presupposes compatibilist freedom (of type 1).

In the light of these distinctions I can state my view about Frankfurt-type examples, which play a prominent role in the contributions of Fischer and Pereboom to the Four Views volume and are discussed by McKenna and Pereboom in what follows. Frankfurt, of course, meant such examples to refute PAP: the principle that an action can be morally responsible only if the agent could have done otherwise. Now my take on Frankfurt examples is different than that of most libertarians. For I agree with Frankfurt, Fischer, Pereboom, McKenna and other defenders of Frankfurt examples that PAP is false. You can see why, since I think there are many acts of free will *of type 2* for which the agents can be morally responsible (even ultimately responsible in the sense of UR) though the agents could not have done otherwise.

My objection to Frankfurt examples is rather that they do not work for acts of free will of type 3, i.e., self-forming acts or SFAs, which, unlike type 2 acts of free will, must be undetermined and such that the agents could have done otherwise voluntarily, intentionally and rationally. Since SFAs are required at some point in our lives if we are to have free will, Frankfurt examples don't show that *free will* (in a UR-entailing sense) does not require alternative possibilities at some point in our lives, *even though they may show PAP is false.*

Consider for example, Pereboom's clever buffered Frankfurt example, Tax Evasion. My argument is not that his agent Joe could not be morally responsible when the controller does not intervene in Tax Evasion because Joe does not attend to his moral reasons. My argument is that the choice Joe makes in Tax Evasion, *whether* he attends to his moral reasons or not, *is not an SFA*. On the one hand, if Joe does *not* attend to his moral reasons, then he will not do otherwise than choose to evade his taxes voluntarily, intentionally and rationally, since he will not attend to any moral reasons for doing otherwise. His choice will not then be an SFA, though it might be a responsible choice. On the other hand, if Joe *does* attend to his moral reasons, the controller would then intervene and prevent him from choosing not to evade his taxes, thus ruling out the possibility of Joe's having a chance to make an SFA he might otherwise have had in the absence of the controller. Either way the choice Joe actually makes, though it might be a responsible choice, is not an SFA.

McKenna gives us a somewhat different example in which an agent's deliberative mechanism is defective, unbeknownst to him, and cannot issue in the choice of one of the options he is deliberating about. This again might be a responsible choice, but is not an SFA, since there is only one robust option available to the agent. The agent may indeed *think* he is making an SFA. But he is not. Free will cannot be guaranteed by introspection, as Mill and others have reminded us. McKenna offers another variation. An agent is deliberating between choices A and B, though unbeknownst to her, B is blocked in some manner and cannot occur. But a third "deliberatively irrelevant" alternative C is nonetheless "causally possible," so the outcome is undetermined. Now C here is either a robust alternative possibility or not? If it is not, if it is something that is merely causally possible and might happen, but is not a choice that might be made voluntarily, intentionally and rationally by the agent, then we do not have an SFA, since there is only one robust alternative

possibility, namely A. If, on the other hand, C *were* a robust alternative possibility (it is hard to see how it could be since it is deliberately irrelevant, but suppose for the sake of argument it was) then we would have an SFA, but then there would also be more than one robust alternative possibility; and the point of the example (like all Frankfurt examples) was to show that A could be a responsible choice, even though the agent lacked any robust alternative possibilities.

To sum up, I think free will entails UR, UR requires SFAs, and the conditions for SFAs are such that they defeat any Frankfurt example. But the very power of SFAs may be their undoing. For the further and deeper problem about free will is how undetermined SFAs are possible at all. How can actions lacking both sufficient causes and motives, and therefore undetermined, be free and responsible actions at all rather than merely matters of luck or chance? This is obviously another matter on which the other authors challenge me. And on this deep question, I believe one must also be willing to think outside the box. What follows is an all too brief sketch of my view about how to do this, followed by answers to some objections raised by the others here and in the book.

First, I believe undetermined self-forming actions or SFAs of type 3 by which we form our character and motives occur at those difficult times of life when we are torn between competing visions of what we should do or become, say, between doing the moral thing or acting from ambition, or between powerful present desires and long term goals, or faced with difficult tasks for which we have aversions. In all such cases, we are faced with competing motivations and have to make an effort to overcome temptation to do something else we also strongly want. There is tension and uncertainty in our minds about what to do at such times, I suggest, that is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium—in short, a kind of "stirring up of chaos" in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. The resulting 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.

When we do decide under such conditions, the outcome is not determined because of the preceding indeterminacy—and yet the outcome can be willed (and hence rational and voluntary) either way owing to the fact that in such self-formation, the agents' prior wills are divided by conflicting motives. An oft-cited example of mine in the book and elsewhere is of a businesswoman on her way to an important meeting who observes an assault in an alley. An inner struggle ensues between her conscience, to stop and seek help, and her career ambitions, which tell her she can't miss this meeting. She has to make an effort of will to overcome the temptation to go on. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not *allow* her effort to succeed. And this is due to the fact that, while she willed to overcome temptation and do the moral thing (help the victim), she also willed to fail, for quite different and incommensurable reasons (her ambitious motives).

Add to this picture now a further point. Just as indeterminism in such cases need not undermine rationality and voluntariness, so indeterminism, of itself, need not undermine control and responsibility. Suppose you are trying to think through a difficult problem, say a mathematical problem, and there is some indeterminacy in

your neural processes complicating the task—a kind of chaotic background. It would be like trying to concentrate and solve a problem, say a mathematical problem, with background noise or distraction. Whether you are going to succeed in solving the problem is uncertain and undetermined because of the distracting neural noise. Yet, if you concentrate and solve the problem nonetheless, we have reason to say you did it and are responsible for it even though it was undetermined whether you would succeed. The indeterministic noise would have been an obstacle that you overcame by your effort.

There are numerous examples supporting this point, where indeterminism functions as an obstacle to success without precluding responsibility—examples first introduced in the 1950s and 1960s by Austin, Anscombe and others.[1] Suppose an assassin attempting to kill the prime minister with a high-powered rifle might miss his target because of some undetermined events in his nervous system that may lead to a wavering of his arm. If he does succeed in hitting his target, despite the indeterminism, can he be held responsible? The answer is clearly yes because he intentionally and voluntarily succeeded in doing what he was *trying* to do—kill the prime minister. Yet his action, killing the prime minister, was undetermined. Another example I've used is of a husband, who, arguing with his wife, swings his arm down on her favorite glass-top table top intending to break it. Suppose that some indeterminism in his outgoing neural pathways makes the momentum of his arm indeterminate so that it is undetermined whether the table will 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 to say to his wife: "Chance did it, not me.")

These examples are not all we want for free will, since they do not amount to genuine exercises of SFAs, like the businesswoman's, where the will is divided between conflicting motives. The woman wants to help the victim, but she also wants to go on to her meeting. By contrast, the assassin's will is not equally divided. He wants to kill the prime minister, but does not also want to fail. (If he fails therefore, it will be *merely* by chance.) Yet such examples provide some clues. To go further, we must dig deeper.

Imagine in cases of inner conflict characteristic of SFAs, like the woman's, that the indeterministic noise which is providing an obstacle to her overcoming temptation is not coming from an external source as in the above cases, but is coming from her own will, since she also deeply desires to do the opposite. Imagine that two crossing (recurrent) neural networks are involved, each influencing the other, and representing her conflicting motivations. (recurrent networks being complex networks of interconnected neurons in the brain circulating impulses in feedback loops generally believed to be involved in higher-level cognitive processing.[2]) The input of one of these networks consists in the woman's reasons for acting morally and stopping to help the victim; the input of the other, her ambitious motives for going on to her meeting.

[1] See Austin (1961) and Anscombe (1971).

[2] Accessible introductions to the role of neural networks (including recurrent networks) in cognitive processing for non-specialists include Churchland (1996) and Manfred Spitzer (1999).

The networks are connected so that the indeterministic noise, which is an obstacle to her making one of the choices, is coming from her desire to make the other, and vice versa—the indeterminism thus arising from a tension-creating conflict in the will, as noted. In these circumstances, when either of the pathways reaches an activation threshold, which amounts to choice, it will thus be like your solving the mathematical problem by overcoming the background noise produced by the other network. And just as when you solved the mathematical problem by overcoming the distracting noise, one can say you did it and are responsible for it, so one can say this as well, I would argue, in the present case, *whichever one is chosen.* The neural pathway through which the woman succeeds in reaching a choice threshold will have overcome the obstacle in the form of indeterministic noise generated by the other pathway.

Note that, under such conditions, the choices either way will not be "inadvertent," "accidental," "capricious," or "merely random," because they will be *willed* by the agents either way when they are made, and done for *reasons* either way—reasons that the agents then and there *endorse*. 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, I argue in the book and elsewhere, rule out each of the reasons we have for saying that agents act, but do not have *control* over their actions, (compulsion, coercion, constraint, inadvertence, accident, control by others, etc.).

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 endeavoring or making an effort to solve *two* cognitive tasks at once, or to complete two competing (deliberative) tasks—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 endeavoring to do. Recall the assassin and the husband. Owing to indeterminacies in their neural pathways, the assassin might miss 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 endeavoring to do. And so it is, I suggest, with self-forming choices, save that in the case of self-forming choices, *whichever way the agents choose* they will have succeeded in doing what they were endeavoring to do because they were endeavoring 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 endeavoring to do two competing things at once in this way, or to solve two cognitive problems at once? Well, we now have evidence 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 endeavoring to resolve plural and competing cognitive tasks. They

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

Such is an all-too-brief sketch of a view. Now some objections. Perhaps the most potent are those based on luck and chance. Indeterminism means: same past, different possible futures. But how is it possible, one might ask, that different actions could arise voluntarily and intentionally from the same past without occurring merely by luck or chance? This question has had a hypnotic effect on those who think about free will. One imagines that if free choices are undetermined, then which one occurs must be like spinning a wheel in one's mind or one must just pop out by chance. No wonder libertarians have looked for some deus ex machina or other to solve the problem, while their opponents have cried magic or mystery. But there is another way to think about the whole matter that I discovered many years ago.

Think, instead, of the indeterminism involved in free choice as an *ingredient* in a larger goal-directed or teleological process or activity, in which the indeterminism functions as a *hindrance* or *obstacle* to the attainment of the goal. If you reflect for a moment, this is what the theory just presented is actually doing. Consider sending a message in Morse Code. The sender taps out the message in dots and dashes. The pulses travel electrically over lines to the receiver where they are reproduced. Now there may be interference due to noise or static in the electrical lines so that the message does not get through, or a distorted message gets through. In that case we have what information theorists call "equivocation" rather than mere noise. But if the message does get through, despite the noise or static, then the goal of the sender is realized. Now if this noise in the lines were the result of indeterminism or chance, whether the message gets through would be undetermined. Yet if the undetermined electrical noise or static was not great enough to cause equivocation, the goal of the process would be realized, despite the interference. The idea then is not to think of the indeterminism involved in free choices as a cause *acting on its own*, but as an ingredient in a larger goal-directed or teleological process or activity in which the indeterminism functions as a hindrance or obstacle to the attainment of the goal. This is the role I have suggested for indeterminism in the efforts preceding undetermined SFAs. These efforts are temporally extended goal-directed activities in which indeterminism is a hindering or interfering element, like the noise or static in the message transmission. The SFAs that result from these temporally extended activities thus do not pop up out of nowhere, even though undetermined. They are the *achievements* of goal-directed activities of the agent that might have failed, but did not.

But can't we say that it is a "matter of chance" whether one of these efforts leading to SFAs succeeds or not? For isn't it true that whether or not an effort succeeds in producing a choice depends on whether certain undetermined neurons

involved in the agent's cognitive processing fire or do not fire (perhaps within a given time frame)? And whether these neurons fire or not is by hypothesis undetermined, is it not, and therefore not under the control of the agent? Well, yes, we *can* say all of these things: And many people reason that this clinches the matter against the claim that the agent is responsible. But they reason too hastily. For the really astonishing thing is that, while all these things can be truly said, *it does not follow* that the agent is not *responsible* for the choice, if the effort succeeds. For, consider the husband swinging his arm down on the table. It is *also* true in his case that whether or not his effort to break the table succeeds "depends" on whether certain neurons in his arm fire or do not fire; and it is *also* true in his case that whether these neurons fire or not is undetermined and therefore not under his control; and we can *also* consequently say in the husband's case that it is a "matter of chance" whether or not he succeeds in breaking the table. Yet, despite all this, it does not follow that he is not responsible for breaking the table, if the effort succeeds. Astonishing indeed! But this is the kind of surprising result one gets when indeterminism or chance plays an interfering or hindering role in larger goal-directed activities, such as efforts to do certain things that may succeed or fail.

It is well to meditate on this: We tend to reason that if an outcome (breaking a table *or* making a choice) depends on whether certain neurons fire or not (in the arm *or* in the brain), then the agent must be able to *make* those neurons fire or not if the agent is to be responsible for the outcome. In other words, we think we have to crawl down to the place where the indeterminism originates (in the individual neurons) and *make* them go one way or the other. We think we have to become originators at the micro-level and tip the balance that chance leaves untipped, if we (and not chance) are to be responsible for the outcome. And we realize, of course, that we can't do that. But we don't have to. It's the wrong place to look. We don't have to micro-manage our individual neurons to perform purposive actions and we do not have such micro-control over our neurons *even when we perform ordinary actions* such as swinging an arm down on a table. What we need when we perform purposive activities, mental *or* physical, is macro-control of processes involving many neurons—processes that may succeed in achieving their goals despite the interfering effects of some recalcitrant neurons. We don't micro-manage our actions by controlling each individual neuron or muscle that might be involved. We don't know enough about neurology or physiology to do that; and it would be counterproductive to try. But that does not prevent us from macro-managing our purposive activities and being responsible when those purposive activities attain their goals.

But does not the presence of indeterminism or chance at least *diminish* the control persons have over their choices or actions? And would that not affect their responsibility? (This is another way in which objections about chance and luck have often been raised.) Is it not the case that the assassin's control over whether the prime minister is killed (his ability to realize his purposes) is lessened by the undetermined impulses in his arm—and so also for the husband and the table? The answer is yes. But the further surprising point is that *diminished control* in such circumstances *does not entail diminished responsibility* when agents succeed in doing what they are trying to do. Ask yourself this question: Is the assassin less

guilty of killing the prime minister, if he succeeds, though he did not have complete control over whether he would succeed or not?

Suppose there were three assassins, each of whom killed a prime minister. Suppose one had a fifty percent chance of succeeding because of the indeterministic wavering of his arm. Another had an eighty percent chance, and the third. a young stud, a hundred percent chance. Is one of these assassins less guilty than the other, *if they all succeed*? Should we say that one deserves a hundred years in jail, the other eighty years, the third fifty years? Absurd. They are all equally guilty if they succeed. The diminished control in the assassins who had an eighty percent or a fifty percent chance does not translate into diminished responsibility when they succeed. Imagine a lawyer for the fifty percent assassin arguing that his client was not guilty because the prime minister's dying as a result of what his client did was a "matter of chance or luck." Not a very telling argument.

The lesson here in summary is that 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. 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. Self-formation is not a gift, but a struggle.

References

Anscombe, G. E. M. (1971). *Causality and determinism.* Cambridge: Cambridge University Press.

Austin, J. L. (1961). Ifs and cans. In J. O. Urmson & G. Warnock (Eds.), *Philosophical papers* (pp. 153–180). Oxford: Clarendon Press.

Churchland, P. M. (1996). *The engine of reason, the seat of the soul.* Cambridge: MIT Press.

Spitzer, M. (1999). *The mind within the net.* Cambridge: MIT Press.