

# Essays in Philosophy

A Biannual Journal

Vol. 10 No. 2

## Book Review

---

*Freedom and Determinism*. Joseph Keim Campbell, Michael O'Rourke, and David Shier, editors. Cambridge, Mass.: MIT Press, 2004. 321 pages, excluding the index. ISBN 0-262-03319-4 (hardcover) \$85.00; ISBN 0-262-53257-3 (paper) \$35.00.

---

This is a wonderful collection of closely reasoned essays by some of the leading contributors to the controversies that constitute the literature of freedom and determinism. Some (e.g., those by van Inwagen, Honderich) are reflective overviews of their own recent work; some (e.g., Nelkin, Arpaly) are significant contributions by younger philosophers. I confess that at least one – the Earman essay – was entirely beyond me (and, unless you are at home with relativity theory and quantum mechanics, will probably be beyond you as well).

The book is divided into three parts. The first, “Determinism, Freedom and Agency,” with essays by Earman, Lehrer, Kane, and Ginet, is devoted to the explication of those three concepts. The second, “The Metaphysics of Moral Responsibility,” is about the relationship of freedom to responsibility; there are essays by Nelkin, Haji, Long, Arpaly, and Fischer. The third, “The Compatibility Problem,” contains essays on some well-know issues concerning the compatibility of free will and determinism, with essays by van Inwagen, Perry, Richard Feldman, Gier and Kjelberg, and Honderich. The editors provide a nice overview in their introduction, and rather than repeat their work I’d like to try to convey some of the flavor of the discussion by looking more closely at several papers that I believe are representative and that hold the most interest for me.

**John Earman, “What We Have Learned and What We Still Don’t Know.”** I said I would pass the Earman paper by and I will, except to make note of an argument that appears to establish the possibility of indeterminism *within* quantum mechanics. The argument is taken from a paper by Laraudogoitia in *Mind*, based upon earlier work by Earman.

I think it is safe to say that for the majority of compatibilists whether the universe is deterministic or indeterministic is not of great relevance; while freedom and responsibility may be compatible with universal determinism they do not require it. On the other hand, libertarians – whether they attribute freedom to alternative possibilities or to “origination” – require the possibility of a break in the causal chain. Kane, as we will see below, is a good example of how indeterminism may be thought to play into responsibility. The purely theoretical possibility of indeterminism within classical mechanics, therefore, is relevant to the discussion.

In this proof of indeterminism we must assume that matter is infinitely divisible; there are no ultimate atoms. If we make that assumption, then it is possible to describe a process in which energy is completely dissipated. The idea is (1) to imagine a line-up of balls of smaller and smaller

size, so that an infinite number of them fit within a given finite space, and (2) to imagine the first and largest of these balls being struck by an equally large ball from outside the given space. Each ball strikes and displaces the next smaller ball, and the entire process is completed within a finite period of time, after which all the balls are again at rest. If this is possible, then given the temporal symmetry of laws in classical mechanics, the reverse procedure is possible as well: A similar sequence of balls at rest may shake off its inertia and begin to move in a direction counter to the one in the original example, finally sending off the intruder ball out of the given space. The process is “self-excited,” as Laraugoitia says, and “is unforeseeable [and presumably without sufficient causal conditions] from the point of view of classical mechanics; it can take place at any instant and it may in fact repeat itself in time any number of times.”

Does this prove the possibility of indeterminism in a way that is useful to the discussion of freedom and responsibility, this most mysterious of all mysterious fields? It does require, as I said, the assumption that matter is infinitely divisible. (Earman remarks: “I see no way to save determinism in this setting. If we cherish determinism, we can only thank the Creator that he did not place us in a world where atomism is false and where Zeno can have his revenge.”) But assume that it does prove the possibility of indeterminism. Although that possibility provides an opening for the libertarian, it doesn’t take him very far. But the proof is fascinating nevertheless.

**Keith Lehrer, “Freedom and the Power of Preference.”** Lehrer had, some years ago, advocated the view that what made action free was that it sprang from the structure of our preferences. It was sort of combination of a Frankfurt-style hierarchical theory, with preferences doing the work that others had used desires for, and a Moore-style conditional analysis. The whole thing was subject to this counterexample, that if your preference structure were imposed on you by someone else, you might satisfy the requirements of the analysis but not be acting freely.

Lehrer’s new response is dazzlingly complex. To the structure of preferences, he adds a “power preference,” which is the actor’s preference for his preference structure. The power preference is self-referential, in that it is not only a preference for the preference structure, it is itself within that structure (and so includes a preference for itself). I think this is an accurate statement of the position: Where an actor would have done A if he (a) had preferred to, and (b) had power-preferred the preference structure that yields that first order preference for A, then doing A was up to him. And if the same thing is true of not-A, then whether or not to do A was up to him, and he acted freely in doing A. The key point is that it is only if I have my preferences because I prefer to have them that I can be said to be acting freely. Furthermore, the reasons for which we prefer A must be reasons we prefer, and there is an “ultrapreference” that serves that purpose. “The causal efficacy of the reasons depends upon my preference for responding in that way.” (p.57) Lehrer does not give us a formal definition that includes both power preferences and ultrapreferences, and I will not attempt to assemble one either.

But what’s to prevent another counterexample of the same sort? The complexity of the analysis, as far as I can tell, does nothing to block the counterexample. *That* work is done by a proviso he calls the “primacy condition,” namely that I have my power preference just because I prefer to have it; “my preference is not manipulated by another.” (p. 54) Lehrer appears to limit the proviso so as to exclude only causation by someone else’s preferences, as in the ‘Braino’ example in which someone has control of my brain in a Frankfurt-like way. He could hardly rule out causation of

preferences altogether, since his aim is to come up with a notion of freedom compatible with causation.

Still, that leaves open the possibility of an example like the following: Suppose Jones is taking part in a laboratory experiment, and must walk around with an electrode implanted in his brain, by means of which the laboratory keeps in touch with him. One day a gust of wind sweeps in through an open window in the lab and knocks some equipment down, causing a bizarre signal to be sent to Jones' electrode. As a result Jones experiences a shift in his preference structure (power preference and all) that leads to a first order preference for dancing on the sidewalk, and he does dance on the sidewalk. It's also true that if he had experienced an alternative shift (or no shift at all), he would not have danced on the sidewalk. Would the preference for dancing be a free preference? Would the dancing be a free action? Had the preference shift been caused by the manipulation of an scheming scientist, I don't think we could avoid the conclusion that the action was not free. Why not the same result here?

I think I know what Lehrer would say: It can't happen – or at the very least it would be the most unlikely combination imaginable. About the possibility of manipulation by the scheming scientist, he had this to say: “[T]here are no such manipulators. Once brainos start getting produced and installed, we shall have a problem about our freedom that we do not have at the moment.” For now, though, nothing to worry about. And the same sort of dismissal might be appropriate for the bizarre case I have imagined.

Perhaps. Discussion of cases of accidents causing sudden changes in motivational structure and of the ensuing question of responsibility are common fare in recent literature; see Tadros, for example, in *Criminal Responsibility*. I'm not sure we can dismiss such questions out of hand.

**Robert Kane, “Agency, Responsibility, and Indeterminism: Reflections on Libertarian Theories of Free Will.”** Kane distinguishes four separate questions: compatibility (whether free will is compatible with determinism); significance (why we should care about free will); intelligibility (whether we can make sense of an incompatibilist notion of free will); and existence (whether such a thing can exist in the natural order). Starting with the compatibility question, he identifies the requirement of alternative possibilities with the requirement that the agent be able to *do* otherwise (if he chooses) (AP), and distinguishes AP from what he calls plurality conditions, which include the ability to *will* otherwise. He concedes that AP and the plurality conditions are not required for every responsible action; but responsible action requires a condition of ultimate responsibility (UR): “[T]o be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient reason (condition, cause, or motive) for the action's occurring.” (pp. 72-73)

UR leads, of course, to infinite regress unless there are some responsible actions for which there is no sufficient reason. Kane finds such actions in self-forming choices (SFAs), choices which are difficult for the agent, which perhaps involve serious moral questions, and which are not causally determined. They are not causally determined because the competition between conflicting motivations occurs at a point at which the causal order of things in the brain is interrupted. These are choices that contribute to the formation of the motivational system that may causally determine our later actions – for which, of course, we will be responsible so long as the causal history can be traced back to some SFA. UR therefore requires, if there are to be any responsible actions, SFAs

for which there were both alternative possibilities and indeterminism.

The problem at this point becomes a problem of intelligibility. How are we to understand “responsible but not causally determined”? Kane is a libertarian, but rejects (for the usual reasons) the Chisholm-Taylor theory of agent causation and every other theory that interjects extra-natural entities into the explanation of action. He calls his own brand of libertarianism teleological intelligibility (TI), insisting that the proper explanation of action is not causal but rather is to be given in terms of reasons or intentions. But how does he avoid calling upon self-moving substances?

The possibility of SFAs depends upon a break in the causal order: There must be choices which are uncaused, but which are yet not a matter of mere chance. Kane finds the possibility of such choices in “microindeterminacies at the neural level.” When an agent is at a crossroads, she is faced with competing possibilities for action, and is highly motivated to choose each of them. These competing motivations may find a window of opportunity in an indeterminacy in the brain, so that neither set of competing motivations is causally determined to succeed. One or the other, nevertheless, must win out. Imagine two sets of motivations, A and B, both strong but each inconsistent with the other. If the agent yields to one, she cannot carry out the other. It might be helpful to imagine that between these motivations and the choice that must be made lies a switch with two settings, 0 and 1. Where the switch will be set is not causally determined, either by a set of prior events or by an agent outside the event-causal order. If the switch happens to be set to 0, the A-motivations will result in a choice to satisfy those motivations, and the B-motivations will be repressed. If it happens to be set to 1, the reverse will happen.

SFAs are therefore not causally determined, but they are not merely random. They are not random because they result from the agent’s motivational system. I have said above that the agent apparently has no control over the indeterministic mechanism that results in one choice or the other. But Kane insists that the agent has control over her action; after all, it resulted from her motivational system, from her values and desires (just as the competing choice would have resulted from her motivational scheme). “In these circumstances, when either of the pathways wins (i.e., reaches an activation threshold, which amount to choice), . . . one can say you did it and are responsible for it.” (p. 80)

Isn’t this just chance, in the end? Kane argues that it is not, and that the idea that it is is due to old habits of thought, habits that must be broken in light of what we know now. I confess that I am skeptical. I admire Kane’s effort to follow out this line of thinking wherever it may lead; but I am, I suppose, in the throes of old ways of thinking. Suppose there are two identical people, each faced with the same difficult choice, one of whom did the right thing and one of whom did the wrong thing, both because of a microindeterminacy in brain function. If I knew all of that, I would not be inclined to praise the one and blame the other. Ultimately, it seems to me, neither is responsible for the outcome. And suppose that as a consequence of a series of such choices, all coincidentally in the wrong direction, one actor now has a bad character, and that the other, as a consequence of a series of choices in the right direction has a good character. Would either one be deserving of praise or blame for their characters?

If Kane were right about this, if (as I understand him) responsibility in this sort of case stems from the fact that the choice is the actor’s choice, meaning only that it comes out of his motivational

system, (remember, in the case of an SFA there is no requirement that the actor be responsible for his motivational system), then compatibilism might well be true, or so it seems to me. All that it takes to make the actor responsible for his action in the critical SFA case is that it is his; that it results from motivations that are his, and not from some external force or some deviant cause. Kane might respond that that is so in general only if the agent is responsible for every sufficient condition for the action, but that response seems to fall apart for his self-forming choices, the motivation for which need not be due to any prior responsible choice.

Kane claims that free will is ultimately an empirical question. There are two problems with this position. The first is that the notion of free will, a common sense notion, may not be finely delineated enough to make it susceptible to empirical testing. The second is that while the truth or falsity of countercausal freedom may be contingent (rather than necessary), it need not be empirically confirmable or refutable. There is one empirical result that would be conclusive as to countercausal freedom, however defined: If there are no microindeterminacies in brain function, libertarianism of whatever stripe will be in a bad way. But whether neuroscience is capable of confirming or refuting that result remains to be seen.

**Dana Nelkin, “The Sense of Freedom.”** Even those who accept determinism may concede that they cannot help feeling free when they make choices. Dana Nelkin sets out to rob this observation of its sting, by showing that the freedom that one must feel in making choices is not a countercausal freedom. She begins her investigation by formulating the claim that the sense of freedom is inescapable something like this:

*1. We have a sense of being free in virtue of being rational deliberators.*

It is clear from her discussion that she will take this to mean:

*2. On each occasion on which we rationally deliberate, we have a sense of being free.*

She then proposes to show that the counter-causal interpretation of “free” in that claim would make the claim false.

*3. On each occasion on which we rationally deliberate, we have a sense of having open alternatives before us.*

She does so by means of a counterexample:

“[I]magine that you know that a brilliant scientist has the ability to fiddle with your brain in a way that causes you to act as she wishes you to. You know that she wants you to vote for Gore over Bush in the upcoming presidential race, and that if you do not decide to vote as she wishes, she will cause you to vote that way. So, for instance, you know that if you were to prepare to vote for Bush or otherwise fail to decide for Gore, the brilliant scientist would cause you to vote for Gore. It seems to me that you could still evaluate the reasons for voting for each candidate and decide to vote for Gore on the basis of those reasons. In this case ... you know that conditions exist which are sufficient for your voting for Gore, while you nevertheless deliberate and decide to vote for him. Further, ... you do not believe you could forgo the action upon which you decide.” (pp.

107-108)

The conclusion of the counterexample is that although you deliberate in this case, it is clear to you that you do not have open alternatives before you. Landmark counterexamples (Gettier, Frankfurt) have a way of attracting silly objections in their early years, and so I offer the following comments with the understanding that they may turn out to be frivolous.

The most obvious question is this: Don't we often deliberate about things not within our power? Might I not consider the reasons for and against being conciliatory with Iran, even though what we do with Iran is not within my power to decide? Might not a high school sophomore deliberate about which candidate would be best to run the country, even though the sophomore cannot vote? So on the face of it, 3. is indeed false, but so are 1. and 2. Who would have thought otherwise?

So in order to take 1., 2., and 3. seriously, something else must be meant by "rational deliberation." I propose the following as roughly what is meant by 3.:

*4. On each occasion on which we rationally deliberate about whether or not to do something X, we have the sense that it is open to us to do X, and also open to us not to do X.*

That is, when deliberation is aimed at issuing in action, we must believe that at least two options are open to us.

If this is the proposition that must be shown to be false, I'm not sure Nelkin succeeds in doing it. Let's look again at the counterexample. There are two possible things Nelkin could have meant here. The alternative to my deliberating and deciding to vote for A might be the scientist causing me *to decide to vote for A*, or it might be the scientist causing me *to vote for A without having decided to*. If the second is the case, then the problem is to understand what it means to cause an action without causing a decision (which Nelkin takes to be the formation of an intention). If it means causing the motions that constitute voting, causing my hand to land on the appropriate lever and make the appropriate movements, then it is not an action at all that is caused. In that case, my choice would be between voting for A and non-action, and the counterexample would collapse, there would in fact be two open alternatives before me. If there is another way to understand how one might cause an action without causing the intention to act, I don't know what it is.

So suppose that what is meant is that the scientist will cause the decision and intention to vote for A. If that is what is meant, then what I know is that wherever my deliberations take me, I will end up deciding to vote for A. Under those circumstances I am not sure that I would agree that I am capable of rationally deliberating about whether to vote for A or B – unless the sort of deliberation I might engage in is the sort of deliberation I engage in when I deliberate about whether the President ought to come to the aid of failing auto makers. In both cases, the deliberation is purely hypothetical, and without effect.

So even if 4. is the claim being tested, I am skeptical about the counterexample. But let's back up a bit. Does anyone really think that the claim at issue here, that rational deliberation bespeaks a belief in freedom, means that whenever we rationally deliberate we must believe that we are free to bring about the result, in any of the senses so far discussed? Each of 2. - 4. appears to have that

implication. But of course we might rationally deliberate about a course of action while believing that *that* course of action might be impossible to bring about, thus perhaps *hoping* that it was up to us to bring it about (or that it was one of the open possibilities) but certainly not *committed to the belief* that it was up to us to bring it about, or that it was one of the open possibilities. I might deliberate about whether to shoot the intruder without being committed to the belief that I could shoot him, either in the sense of having it as an open alternative before me or in the sense that it was somehow up to me. I might not be a good enough shot to be sure of being able to do it; I might not have the psychological fortitude to do it. (And I take it my deliberation would not just be about whether to try to shoot him, but about whether it would be right for me to shoot him.) So on at least some occasions there must be rational deliberation about a state of affairs without a belief that it is up to us to bring about that state of affairs.

The claim in question, if it is to survive this simple criticism, cannot really be that on each occasion of rational deliberation about an action there was a commitment to a belief in freedom to perform that action. A different way of understanding the claim would be this: That there would have been no occasion to engage in practical reasoning if we believed from the beginning (if such a thing were possible) that all of our decisions were caused by factors beyond our control. The faculty or practice of deliberation would never have had a reason to develop. If that is so, then the belief that we are free in the sense of having open alternatives in at least some cases would be a necessary condition of engaging in the practice of rational deliberation.

The claim that we started with was that rational deliberation requires a sense of freedom. Call any interpretation of that claim that requires a sense of freedom for each action that we deliberate about a limited interpretation, and an interpretation that requires only that we have the sense that we are at least sometimes free a general interpretation. Since Nelkin's counterexample to the limited countercausal interpretation of the claim is in a way the crux of her argument, it would be helpful to have some of the questions I have raised answered.

**John Martin Fischer, "The Transfer of Nonresponsibility."** As we have seen, Robert Kane premised his argument in part on the very natural intuition that if our actions are caused by our beliefs and desires, we are responsible for those actions only if we are responsible for those beliefs and desires. This leads to an infinite regress unless there are actions that (a) are not caused by prior beliefs and desires, and (b) are themselves partial causes of our later beliefs and desires. Kane sought to end the regress by positing the existence of SFAs.

Some time ago Peter van Inwagen attempted to formalize the underlying intuition into a proof of incompatibilism. The proof went (very roughly) like this: If determinism is true, then any future action of mine will follow logically from the conjunction of past events about which I had no choice and laws of nature about which I have no choice. But if I have no choice about A and no choice about B, and if the conjunction of A and B logically entails C, then I have no choice about C. (This last proposition, which attempts to formalize the underlying intuition, is the one version of what is known as the *transfer* principle, in this case logically transferring powerlessness with respect to the past and to laws of nature to human action.) Thus if determinism is true, I have no freedom (understood countercausally) as to any future action; and so determinism and freedom are incompatible.

You may or may not accept the argument. Van Inwagen himself, in his contribution to this volume,

considers objections to the principle. But even if you accept it, one of the things that appears to follow from the Frankfurt counterexamples is that a lack of freedom (understood countercausally) need not entail a lack of responsibility, and so the incompatibility of *freedom* and determinism need not mean the incompatibility of *responsibility* and determinism. That is, while the absence of open alternatives may transfer from the past to the present, nonresponsibility does not. Fischer and others who accept the first sort of incompatibilism but reject the second consider themselves to be *semi-compatibilists*.

In this paper, Fischer picks up the debate he has been having with advocates of a transfer-of-nonresponsibility principle. Fischer and Mark Ravizza have formulated counterexamples to a series of formulations of this principle. The counterexamples depend upon the possibility of overdetermination: If I (intuitively freely) cause something (*q*, say) to happen, but *q* is also caused by some natural event, then I remain responsible for what happened even though (1) something happened in the past (*p*) that I was not responsible for; and (2) that something, *p*, was a sufficient causal condition for *q*, and I was not responsible for that causal fact (nor was anyone else). For example, suppose I set fire to some brush, intending that the fire reach Harry's cabin and burn it down; and unknown to me a natural brush fire has started on the other side of the cabin, and the two fires reach the cabin at the same time, and either would have been sufficient to burn it down. Then I remain responsible though I am not (as the lawyers say) the "but-for" cause of the house burning down. It would have burned down in any case.

Critics have argued that though the counterexample shows that responsibility for an outcome *q* is possible even though there is a fixed past event and a law of nature connecting that past event with *q*, it only works where there are two simultaneous paths to *q*, in the example above the fire I set and the natural fire. Where there is only a single line of causation leading to *q*, if *q* is the causal outcome of past events I am not responsible for then I am not responsible for *q*. Fischer has two responses to this. The first is that the argument was intended to show that determinism was incompatible with responsibility and it does, because the counterexample shows that there are cases in which the premises of the argument are true but the conclusion is false. So the transfer principle is false, even if the compatibilist is left in a bad way in connection with the one-path cases. Still, there are good reasons for accepting compatibilism, and it requires a good argument against to counter those reasons.

Fischer's second response is his familiar claim that the most that has been shown is that there is a dialectical stalemate. In the course of setting this response, Fischer makes the following argument: van Inwagen and others have cited certain sorts of examples in support of the transfer principle. For example,

[S]uppose that an earthquake takes place in the middle of the Pacific Ocean, and that no one is even partly morally responsible for this. Also suppose that if this earthquake occurs, then a tsunami will hit the coast of California, and no one is even partly responsible for this fact. It appears to follow that a tsunami will hit the coast of California, and no one is even partly morally responsible for this. (p. 192)

The example doesn't involve a human actor, of course, but it is supposed to illustrate the intuitive nature of the principle; that is, it is supposed to show that we accept that the conclusion follows from the premisses. But a weaker principle is available, Fischer says, that will explain such

examples without supporting the conclusion of incompatibility of determinism and responsibility. The principle is this:

If (1)  $p$  obtains and no one is even partly morally responsible for  $p$ ; and (2) if  $p$  obtains, then  $q$  obtains, and no one is even partly morally responsible for the fact that if  $p$  obtains, then  $q$  obtains; and (3) on the actual path that leads from  $p$ 's obtaining to  $q$ 's obtaining, either there is no factor that at least prima facie could be thought to ground moral responsibility, or there is some factor that uncontroversially undermines moral responsibility . . . .; then (4)  $q$  obtains, and no one is even partly morally responsible for  $q$ . (pp. 201-202)

In other words, just as the supporters of the transfer principle inserted a provision into that principle that excluded the possibility of two-path examples, Fischer has constructed a principle that excludes the possibility of human causes that meet a compatibilist definition of responsibility. This principle explains the uncontroversial examples called upon to support the transfer principle of course, since none of those examples involve human actors.

But if this sort of ad hoc fiddling is what is required to make progress in this area, then Fischer appears to be right. Among all the debates that surround the issue of freedom and determinism, this particular debate does seem to be at a stalemate.

Michael Corrado  
University of North Carolina at Chapel Hill