



Review: [untitled]

Author(s): John Martin Fischer

Source: *Noûs*, Vol. 33, No. 1 (Mar., 1999), pp. 133-143

Published by: [Blackwell Publishing](#)

Stable URL: <http://www.jstor.org/stable/2672019>

Accessed: 24/02/2011 13:21

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing is collaborating with JSTOR to digitize, preserve and extend access to *Noûs*.

<http://www.jstor.org>

Alfred R. Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York and Oxford: Oxford University Press, 1995), viii + 271 pp.

JOHN MARTIN FISCHER
University of California, Riverside

Autonomous Agents is a very fine book. The book is meticulously argued, and it is written carefully and clearly. It contains highly illuminating discussions of many of the central notions pertinent to self-control (and weakness of the will) and autonomy. It would be extremely useful for anyone interested in action theory, free will, moral responsibility, and autonomy. I shall begin with a brief summary of the book. Then I shall highlight what I take to be some of the most interesting components of the book; these are often extended critical discussions of influential arguments that have been offered by other philosophers. The careful presentation and analysis of these arguments is one of the most salient strengths of this book. Finally, I shall offer some of my reactions to some aspects of the book.

The book is divided into two parts. Part I develops an account of self-control. Mele argues that even an ideally self-controlled person may lack autonomy. Part II gives an account of autonomy and explains what must be added to self-control to get autonomy. The relevant notion of self-control, for Mele, is the “contrary of *akrasia*” (weakness of the will or incontinence). (p. 3) And the relevant notion of autonomy is, to a first approximation, an actual condition of an agent exhibiting “self-rule” or “self-government.” (p. 3)

A kind of *akrasia* that, as Mele points out, has “received pride of place in the literature,” is uncompelled intentional action that the agent recognizes to be contrary to a decisive better judgment that he consciously holds. (p. 16) An interesting feature of Mele’s discussion of *akrasia* is his contention that there are various other sorts of *akrasia*, including akratic acts that *conform* to one’s best judgment. (I’ll return to this possibility a bit later.)

On Mele’s picture of practical reasoning, best judgments—an agent’s judgment about what it is best for him to do—typically lead to intention formation via what Mele calls a “default procedure.” But given human nature, a best judgment does not ensure the formation of a corresponding intention—best judgments must compete with conflicting desires in certain contexts.

Decisive best judgments, then, are subject to defeat by opposing motivation. Also, they are supportable by exercises of self-control. But how does self-control work? That is, if an individual judges it best to *A* but wants more to *B*, where *A* and *B* are incompatible acts,

how exactly can an exercise of self-control be motivationally possible for the individual? Mele points out that in a context such as the above, an agent may want to *B* more than to *A*, but also want to *A* and *C* more than to *B*. Thus, he may coherently be seen to act in accordance with a strongest desire in exercising self-control (in resisting *B*). For example:

a man with limited funds may want to see a certain expensive play tonight more than he wants to see a movie tonight, but want to see a movie and eat dinner at a restaurant tonight more than he wants to see the play. (He may realize that, given his financial situation, he cannot see the play and eat dinner at a restaurant. (p. 42)

Mele goes on to show how various strategies for exercising self-control exploit the above structure. This is one of several places in the book where Mele employs to excellent effect his command of literature in related fields, particularly psychology. (Chapter Three) Mele also devotes a considerable section of Chapter Three to showing how self-control is possible even when the above structure does not apply.

A particularly interesting feature of Mele's discussion of *akrasia* is his contention that one can manifest *akrasia* by performing an act that coincides with one's best judgment. Similarly, one can exercise self-control on behalf of an act that goes against one's best judgment. (Thus, the idea that akratic action is uncompelled, intentional action that conflicts with one's best judgment is merely a first approximation to the truth.) Here is Mele's example:

Young Bruce has decided to join some wayward Cub Scouts in breaking into a neighbor's house, even though he decisively judges it best not to do so. Suppose that at the last minute Bruce refuses to enter the house and leaves the scene of the crime. His doing so because his decisive judgment has prevailed is one thing; his refusing to break in owing simply to a failure of nerve is another. In the latter event, Bruce arguably has exhibited weakness of will: he 'chickened out,' as children are wont to exclaim. (p. 60)

Mele also points to an alternative version of the case in which Bruce experiences some trepidation about the housebreaking and tries to steel himself for the deed. Although he judges it best not to participate in the crime, he successfully masters his fear, and he proceeds to pick the lock. Here Bruce has exhibited strength of will; some exercises of self-control are not performed in the service of a best judgment. (p. 60)

If *akrasia* can be exhibited in cases of action in accordance with one's best judgment, what is then common to all cases of *akrasia*? According to Mele, all akratic action is action in which a practical commitment is thwarted by noncompelling competing motivation. (p. 74) When the akratic action is against a decisive better judgment, the relevant practical commitment is an evaluative commitment—an answer to the question of what it is best to do. When the akratic action accords with a decisive better judgment, the relevant practical commitment is an executive commitment—an answer to the question of what to do.

Mele plausibly contends that the spheres of *akrasia* and self-control extend beyond actions to beliefs and feelings (and emotions). Mele's discussions here again exhibit his considerable knowledge of relevant psychological literature.

In Chapter Seven there is a very insightful discussion of an individual who is extremely compulsive about his work. This sort of compulsion imposes rigid constraints and routines on the individual, and cuts him off from prospects for a more satisfying life. Interestingly,

the victimization here is *self-victimization*. Thus, if one is inclined to think that such an individual may lack autonomy, autonomy would not simply be a matter of “self-rule” or “self-government.” As Mele puts it, “certain *kinds* of self-rule or self-government...are self-limiting, self-oppressing, and self-victimizing in a way that is at odds with a robust autonomy.” (p. 126)

Mele responds to such examples in a careful way. He does not insist that compulsive individuals are non-autonomous. Rather, he suggests that one element of a *sufficient* condition for the possession of personal autonomy is that an agent’s life be “unsullied by self-oppression or self-victimization.” (p. 126) He uses the tag, “continuous mental health,” to refer to this condition of non-self-victimization.

Mele considers, at the end of Part I, an “ideally self-controlled person.” (pp. 121ff) Such a person does not exhibit any degree of *akrasia* with respect to his actions, beliefs, and emotions. Now Mele asks whether such an imaginary individual is *autonomous* (simply in virtue of being ideally self-controlled). The answer, according to Mele, is no. This is because an autonomous agent must possess the capacity to reflect critically upon his preferences and desires, and the ability either to identify with these or to change them in light of higher-order preferences and values. (Mele here follows Gerald Dworkin¹: p. 122) Mele now points out that even an ideally self-controlled person might have every process of critical reflection regulated or guided by principles or values that are the products of brainwashing or other forms of “mind control.” (p. 122) He thus concludes that even an ideally self-controlled person may not be *autonomous*.

In Part II Mele turns to the question of what autonomy consists in (and thus, the question of what must be added to self-control to get autonomy). In an important discussion, he forcefully argues that autonomy is not simply a matter of the internal arrangement of an agent’s psychological states (and the capacities and dispositions the agent possesses at a time). Mele thus argues for a kind of “externalism” about autonomy, rather than “internalism.” Specifically, autonomy is a *historical* notion, according to which the history behind an agent’s configuration of internal motivational states is crucial to the question of whether he is autonomous. Mele points out that two agents can be type-identical with respect to their internal psychological states (and capacities, dispositions, and so forth) and yet can be very different with respect to their autonomy, given the different histories behind the “time-slice” features of their mental states. So, for example, if one agent has been brainwashed or significantly manipulated and the other has not, this can make a crucial difference to their autonomy.

Autonomy is in this respect similar to other historical notions. Mele’s examples are sunburns and genuine currency. A sunburn is not simply a certain condition of the skin; it is that sort of condition *caused by excessive exposure to the sun*. Similarly, something’s being genuine currency—and not counterfeit—is not simply a matter of how it appears at a given time; it is partly a matter of its causal history. So, also, with autonomy.²

On Mele’s account, it is a necessary condition of a pro-attitude’s being possessed autonomously that it be possessed *authentically*. Authenticity is history-sensitive. As Mele puts it, “A necessary condition of an agent *S*’s *authentically* possessing a pro-attitude *P* (e.g., a value or preference) that he has over an interval *t* is that it be false that *S*’s having *P* over that interval is, as I will say, *compelled**—where *compulsion** is *compulsion not arranged by S*.” (p. 166) Further, Mele offers the following sufficient condition for “*compulsion**”:

2*. If an agent *S* comes to possess a pro-attitude *P* in a way that bypasses *S*’s (perhaps relatively modest) capacities for control over his mental life; and the bypassing issues

in *S*'s being practically unable to shed *P*; and the bypassing was not itself arranged (or performed) by *S*; and *S* neither presently possesses nor earlier possessed pro-attitudes that would support his identifying with *P*, with the exception of pro-attitudes that are themselves practically unsheddable products of unsolicited bypassing; then *S* is compelled* to possess *P*. (P. 172)

We have now what Mele characterizes as a “partial” answer to the question of what must be added to ideal self-control to get autonomy: a certain sort of authenticity. Now Mele develops a bifurcated discussion. First, he assumes (for the sake of the discussion) that compatibilism about causal determinism and autonomy is true. He then develops a set of sufficient conditions for autonomy, given this assumption. Second, Mele assumes that incompatibilism about causal determinism and autonomy is true, and he develops a set of sufficient conditions for incompatibilistic autonomy. His official view is *agnostic* with respect to the issue of whether compatibilism or incompatibilism about causal determinism and autonomy is true. He does not take a stand as to whether causal determinism or indeterminism obtains, or whether compatibilism or incompatibilism is true.

Assuming compatibilism and that the relevant agent is ideally self-controlled and is mentally healthy, Mele contends that we have a psychologically autonomous agent if we add the following three conditions:

1. The agent has no compelled* motivational states, nor any coercively produced motivational states.
2. The agent's beliefs are conducive to informed deliberation about all matters that concern him.
3. The agent is a reliable deliberator. (p. 187)

Each condition is elaborated in light of various examples of problematic manipulations and interferences.

Robert Kane has argued that no compatibilist account can adequately address the possibility of certain sorts of manipulation.³ More specifically, Kane has contended that compatibilism cannot give a good response to the possibility of “CNC control”—“covert, non-constraining control.” Mele's conditions are in part an effort to show that a compatibilist indeed has the resources to reply to Kane's worries. Mele discusses various cases of CNC control, and argues that his conditions provide a good reply to Kane's challenge to compatibilism.

Mele now takes the incompatibilistic path. That is, he develops an account of autonomy, assuming that incompatibilism about causal determinism and autonomy obtains. Basically, the account consists of the sufficient conditions for compatibilistic autonomy embellished by the requirement that there be a certain sort of indeterminism. But it is a notorious problem for the libertarian to say exactly where the indeterministic gap should be installed.

Think of practical reasoning roughly as follows. One has a number of standing desires, preferences, and values. These are inputs into one's deliberations, together with various pieces of information (and general theories and background presuppositions). One combines and weighs these inputs, on the basis of which one forms a decisive best judgment. If things go well, one forms an intention that is in accord with the decisive best judgment and acts in accordance with the (proximal) intention.

Prior to forming one's decisive best judgment, a number of considerations will come to mind: one will have a sequence of occurrent beliefs. Mele argues that the most felicitous place for the indeterministic gaps are at certain points (but not necessarily all points) in the sequence of occurrent beliefs prior to the formation of the decisive best judgment. Placing the gaps anywhere further along the chain that leads to action would seem to introduce unacceptable arbitrariness and an associated lack of control. Further, Mele argues that placing the gaps in the internal sequence of beliefs need not issue in a diminution of control. The idea is that it is not determined precisely which thoughts one will have and in precisely which order, but the general nature of the thoughts will be constrained by one's standing dispositions, preferences, and values. The view here is similar to that suggested by Daniel Dennett in his essay, "On Giving Libertarians What They Say They Want."⁴

Having given both a compatibilist and an incompatibilistic set of sufficient conditions for autonomy, Mele ends by highlighting the salient advantages and disadvantages of the various available positions. Compatibilistic autonomy is attractive because it can help to render secure our belief in ourselves as autonomous agents. Also, it can (according to Mele) answer the challenge posed by Kane (to address various cases involving sophisticated covert manipulation). But the disadvantage is that it cannot account for the view that an agent cannot be autonomous if each internal state or event, and each action, is the product of some deterministic causal chain. (p. 252). Whereas it is not obviously correct that each state's being the product of a deterministic causal chain rules out autonomy, it is not unreasonable for someone to think that it does.

An apparent advantage of the sort of libertarianism developed by Mele (which involves the requirement of "internal" or "doxastic" indeterminacy) is that it can accommodate the deep intuition just mentioned—that an agent is not autonomous if his states and actions are the result of a deterministic causal chain—consistently with the agent's having genuine control. (The indeterministic gaps are allegedly not placed in locations which would erode control.) A major problem, however, with this view, even if we accept that doxastic indeterminism solves the "control problem" (more on this below), is that for all we know, we are internally deterministic (or internally indeterministic only in ways of no theoretical use to libertarians). (p. 253). Thus, on this approach, we cannot be confident that we are in fact autonomous.

Agnostic (about compatibilism) belief in autonomy has certain advantages. According to Mele, it lacks the disadvantages of compatibilism and incompatibilism. That is, it need not argue that autonomy is in fact compatible with determinism nor that we are in fact internally indeterministic in the envisaged way (*i.e.*, doxastically indeterministic). Unfortunately, this view leaves it open that (1) no set of compatibilistic sufficient conditions is sufficient for genuine autonomy, and (2) we are not internally indeterministic in the relevant way. Thus, the agnostic position seems to involve a belief in autonomy that cannot be decisively supported. (I shall return to this point below.)

Finally, Mele rejects the doctrine he calls "nonautonomism," which is the view that no human being is autonomous. Mele claims that nonautonomism has decisive disadvantages. He says:

[Nonautonomism} must provide a convincing argument for incompatibilism and convincing grounds for the thesis that human beings are not internally indeterministic in a way required for incompatibilist autonomy. Compatibilist and agnostic believers in human autonomy take comfort in the absence of any convincing argument of the first sort. ... Further, libertarians and agnostic autonomists rightly observe that, given our present knowledge and ignorance about the workings of the human organism, we

lack grounds for confidence that internal indeterminism of a sort required by a coherent, physicalistically minded libertarianism is not a feature of actual human beings. (p. 254)

In the end, then, Mele accepts agnostic autonomism (on the basis of his assessment of the advantages and disadvantages of the various views). This combines agnosticism about compatibilism and libertarianism with a belief in autonomy. The agnostic autonomist does not commit himself to the contention that compatibilist autonomy is genuine autonomy or that incompatibilist autonomy is genuine autonomy; nevertheless, he does contend that we are genuinely autonomous. One way of putting this sort of view is as follows: the agnostic autonomist may believe that if the relevant sort of internal indeterminism does not actually obtain, then compatibilism is true.

II. Some Noteworthy Discussions

One of the striking merits of Mele's book is the excellent critical analyses of various important and influential arguments in action theory. I want to call attention to some of these very insightful and useful critical discussions.

Various philosophers have supposed that one could not offer an adequate analysis of the phenomena pertinent to *akrasia* without positing "higher-order" preferences or desires. Although this is perhaps an attractive and natural supposition, Mele argues forcefully that one can adequately analyze these phenomena without adverting to the apparatus of higher-order motivational states.

It is tempting to think of akratic action as action against a higher-order preference—a preference not to act on the actually-motivating desire. And there is no doubt that this is one way of beginning to understand *akrasia*. But Mele points out that the apparatus of practical commitments (evaluative and executive) can be employed to give a more nuanced and incisive analysis of *akrasia*; once this apparatus is in place, higher-order preferences seem otiose. (p. 79)

Gary Watson has presented an influential and challenging argument for skepticism about the possibility of akratic action.⁵ Watson argues that *akrasia* cannot be explained in terms of an agent's choosing not to resist the relevant desire or by his making a culpably insufficient effort to resist. The agent's choosing not to resist, according to Watson, would be to change his best judgment—and thus it could not explain action *against* his best judgment. Similarly, an insufficient effort cannot be due to a judgment that the effort is not worthwhile, because the judgment that it is worth the trouble is implicit in the relevant best judgment. (p. 107)

Mele responds to Watson with a subtle analysis that carefully distinguishes between decisive better judgments and choices. As Mele illustrates with an example, "judgments are one thing and choices another." (p. 107) He also points out that insufficient effort need not be due either to a judgment that taking preventive measures is not worth the trouble (or to a misjudgment about the amount or kind of effort required). (p. 107) Anyone interested in Watson's important challenge to the possibility of weakness of will should consider Mele's thoughtful reply.

There are various other useful discussions which I can only briefly refer to here. Peter van Inwagen presents a fascinating "slippery slope" argument that is designed to show (contrary to his own views, of course) that genuine action and freedom require causal determinism.⁶ Mele calls van Inwagen's argument the "Piano argument." (p. 197) The

argument begins with a scenario in which a demon has a piano keyboard made of “subtle matter” that is connected by a “subtle wire” to a certain thief’s brain; by what the demon plays, he directs the motions of the atoms in the thief’s brain and can thereby direct the thief’s inner life, including the thief’s deliberations. Clearly, under these circumstances the thief does not freely rob the poor-box of the church. There follows a series of modifications of the example which lead ultimately to the thief’s stealing from the poor-box via an *indeterministic* process. The challenge for the libertarian is to say exactly where in the sequence we move from a lack of freedom to the possession of freedom. Mele gives a sustained, subtle, and insightful analysis of van Inwagen’s discussion. (pp. 196–204)⁷

Another noteworthy discussion is Mele’s analysis of Robert Kane’s libertarian picture of deliberation and freedom.⁸ Unlike Mele’s libertarian picture, on Kane’s view there is an indeterministic gap between prior character and motives on the one hand and choices on the other. On Kane’s account, in cases of conflict an agent may make an effort to resist a particular desire; it is undetermined whether the effort will be successful. Kane emphasizes an analogy with quantum mechanics: just as it is undetermined whether an isolated particle moving toward a thin atomic barrier will penetrate the barrier, it can be undetermined whether an effort to choose a particular course of action will in fact succeed.

Mele contends that Kane’s placement of the indeterministic gap between deliberation and choice raises significant questions about the agent’s possession of genuine control. (p. 206) There is also a nice discussion of certain other problems with Kane’s overall view. For example, Kane appears to be committed to the view that freedom is only exhibited in contexts of conflicting motivations, which seems intuitively incorrect. Also, Mele shows how Kane’s view appears to rule out the possibility of genuinely akratic actions, which, again, seems seriously problematic from an intuitive point of view. (pp. 207–208)

Galen Strawson has argued that the common-sense notion of libertarian freedom is internally incoherent insofar as it is implicitly committed to a problematic infinite regress of choices.⁹ Mele gives a careful and convincing refutation of Strawson’s argument. (pp. 221–227)

Finally, Richard Double has also argued that free will (and presumably autonomy) is an internally inconsistent concept.¹⁰ He proceeds by setting out six conditions that he contends are both required by common-sense understandings of free will and jointly impossible to satisfy.¹¹ Mele argues against Double’s pessimistic conclusion about free will and autonomy. (pp. 241–243)

Anyone interested in the collection of issues related to action, intention, weakness of will, free will, moral responsibility, and autonomy will benefit from Mele’s careful, subtle, and incisive discussions.

III. Two Areas of Concern

I am faced with the difficult task of doing a critical notice of a book, with almost all of which I agree! I have however managed to find two areas in which I have substantial questions about Mele’s arguments. I do not necessarily find Mele’s treatments in these areas unacceptable, but at least I have some considerable doubts and questions.

III.1. The first area of concern is Mele’s development of a libertarian view on which the indeterministic gap is placed within the sequence of doxastic states constitutive of deliberation. (Of course, Mele is officially agnostic about the truth of libertarianism, but he defends this version as the most plausible form of libertarianism—one which solves the “control” problem.) Although I myself have presented something like this as a reasonable

version of libertarianism, I have some doubts as to whether it does indeed solve the control problem.¹²

On Mele's picture of libertarian agency and control, the precise sequence of doxastic states leading ultimately to a judgment as to what is best is undetermined, although the agent's standing desires, values, and general dispositions constrain the content and presumably the order (to some degree) of the states. Everything else in the sequence—the formation of a best judgment based on one's deliberations, the transitions to an intention, a proximal intention, and an action are all deterministic. Whereas Daniel Dennett has suggested a similar picture, a distinctive twist is added by Mele: he argues that indeterminism of the sort he posits—internal, doxastic indeterminacy—is no worse, in respect to control, than determination in this portion of the sequence leading ultimately to action. This is because the agent typically does not directly control the sequence of doxastic states he undergoes—this is generally a passive rather than an active process. Mele says:

...notice that we are not always in (proximal) control of which of our beliefs come to mind anyway, even if determinism is true. Assuming determinism, everything that happens on this front is causally determined, but the causal story often does not place the agent in the driver's seat. (p. 215)

So Mele's suggestion is that, even though installing indeterminacy may seem to erode genuine control, installing it at this particular place (at some points in the sequence of doxastic states which constitute deliberation) does *not* diminish control that otherwise would be present (under the assumption of determinism); a proponent of doxastic indeterminacy is thus not worse off than a proponent of doxastic determination, with respect to control.

Additionally, Mele contends that by installing indeterminacy in the sequence leading to action, one is able to preserve the crucial libertarian belief in alternative possibilities or freedom to choose and do otherwise. Of course, one has to be careful about the specification of the temporal index associated with the alternative possibility. It will be true during the period of doxastic indeterminacy in the agent's deliberation that he has the power to form different best judgments and thus to pursue more than one path in the future; but after the period of doxastic indeterminacy (and just prior to the formation of the agent's best judgment), he will no longer have alternative possibilities, insofar as the process is now causally deterministic.

My puzzle could be put as follows. How can adding arbitrariness of the sort envisaged—the lack of determination of the beliefs that come to mind during deliberation—to a causally deterministic process yield genuine control? A libertarian of course will contend that an *entirely* deterministic process does not contain genuine control by the relevant agent. How, then, can installing the sort of indeterminacy envisaged—indeterminacy as to which belief states will come to the agent's mind—transform the sequence from one of lack of control to one containing control? This smacks of alchemy.¹³

Perhaps my point could be made in terms of a crucial distinction (of which Mele is certainly aware) between an agent's having control over what happens, and its being the case merely that something different might have occurred. If an agent has genuine control in the sense of possessing alternative possibilities, he can make it the case that one path is followed, or another path is followed, *in accordance with what he judges best and chooses*. He can deliberately pursue one course of action, or deliberately pursue another; what path the world takes (at least in certain respects) is "up to him." In contrast, when it is merely possible that something different have occurred, the path the world takes need not depend in the relevant way on the agent. In a genuinely random event, presumably there are

various metaphysically open possibilities; but by definition no agent has control over what happens.

Now it seems to me that, whereas it may well be possible that Mele's libertarian agent do something different from what he actually does, it is not clear that he has genuine *control* over what he does. Mele admits that the precise sequence of doxastic states can have an effect on what the agent judges best (and then does); given that the sequence is not (entirely) determined by prior states of the agent (although it is constrained by such states), it is not clear that what the agent judges best and then does is genuinely up to him.

Mele's point that even on a deterministic model, the agent is not "in the driver's seat" with respect to which considerations come to mind (and thus that the libertarian is here not losing anything with respect to control) is an intriguing and suggestive idea. But presumably the compatibilist will point out that, even though the agent does not directly control what belief-states come to mind (in the sense of choosing them or willing them), they are envisaged as strongly connected to the agent's prior states to the extent that they are a *deterministic product* of those past states. Under determinism, one's prior states—desires, beliefs, values, general dispositions—*determine* the precise content and ordering of the subsequent doxastic states (that constitute deliberation), even if the agent does not directly *control* what doxastic states he will be in (and thus is not in the "driver's seat," in this sense).

It may then be possible to argue that one *does* give up some measure of control, when one shifts from thinking of the doxastic sequence as deterministic to thinking of it as indeterministic: one gives up the notion that the states constituting one's deliberations are an "outflowing" of the agent's prior states in a strong sense. How one assesses Mele's libertarianism seems to me to hinge on whether one believes that the doxastic states that constitute one's deliberations can be a genuine "outflowing" of the agent's prior states, even though they are not causally determined by those prior states. I am not convinced, however, that from the mere fact that the agent is deemed passive in regard to his doxastic states even under determinism, it follows that one does not attenuate the agent's control in positing doxastic indeterminism. Thus, whereas Mele's suggestion on behalf of the libertarian is suggestive and argued with considerable resourcefulness, I am left with nagging doubts about the strategy.

III.2. I also have some questions about Mele's preferred position, agnostic (about compatibilism) autonomism. It is somewhat perplexing to me why Mele opts for this position, despite Mele's spirited explanation and defense of it. It seems to me that the natural conclusion to draw from agnosticism about compatibilism and libertarianism would be *agnosticism* about autonomy; but Mele wishes to combine agnosticism about compatibilism and libertarianism with a belief in *autonomy*.

Think of it this way. Presumably Mele is an agnostic about compatibilism and also about libertarianism because, although he recognizes some reasons to adopt these views, he does not think that there are *knockdown* reasons to accept these views. So, in the absence of knockdown reasons for (say) compatibilism, Mele opts for agnosticism about compatibilism. Why, then, change the standards when thinking about autonomy? Mele does *not* think that there are knockdown reasons to believe in autonomy; he says:

The doubly agnostic autonomism that I have described officially leaves it epistemically open (1) that no set of sufficient conditions for *compatibilist* autonomy is sufficient for *autonomy* and (2) that human beings are not internally indeterministic in a way theoretically useful to incompatibilist believers in autonomy. In leaving this open, agnostic autonomism leaves it epistemically open that no human being is autonomous. (p. 253)

Why does Mele believe in autonomy in the absence of a knockdown argument, but fail to believe in compatibilism in the absence of a knockdown argument? It seems that, given the standards he accepts in assessing compatibilism, he should be an *agnostic* with respect to autonomy.¹⁴

Now someone may believe that there are strong reasons to accept a belief in autonomy which, although they fall short of being decisive, nevertheless warrant a belief in autonomy. Further, he may believe that there are no similarly strong reasons to accept a belief in compatibilism. But I am not confident that this is indeed Mele's view. And if it is, I would be interested in a more explicit development of this view.

IV. Conclusion

This book is sensible, clear, and meticulous throughout. Mele's discussions of *akrasia* are as subtle and insightful as one could hope for. Additionally, his analyses of the various issues relating to autonomy are penetrating and helpful. Mele combines rigorous, sophisticated philosophical argumentation with an encyclopedic command of relevant literature in related fields (such as psychology). I appreciate Mele's straightforward (if sometimes somewhat intricate) philosophical style, and his intellectual honesty.

Notes

¹Dworkin 1988.

²I (and my co-author) have made a similar argument with respect to moral responsibility in Fischer and Ravizza 1994 and 1998, esp. pp. 170–207.

³For a recent development of this view, see Robert Kane 1996.

⁴Dennett 1978.

⁵Watson 1977.

⁶van Inwagen 1983, pp. 126–142.

⁷I have for a long time been baffled by what van Inwagen says about the slippery slope argument:

At a certain point in the story, the reader will remember, the freakish demon and his keyboard of subtle matter have been reduced to a wire-shaped thing of brain-stuff. The next step is to suppose that this thing is a *natural part* of the thief. This next step is, in my view, the one that makes the difference. ...the *Mind* argument [the slippery slope argument in service of the position, often found in the pages of *Mind* (in a certain era), that causal determinism is a necessary condition of freedom and moral responsibility]...errs in supposing that the step from 'wire-shaped thing' to *natural wire-shaped thing*' is an obviously harmless step. (van Inwagen, 1983, p. 135.)

I do not see how it can make *any* difference whether the part of the brain in question is a "natural part" or not, as long as it is appropriately integrated into the functional economy of the human being's brain. Surely, if the thief were to raise his arm as a result of the ordinary sort of practical reflection and deliberation, this would be no less an action (or behavior under his control) if it were a prosthetic arm. Similarly, it would seem to make no difference whether the relevant part of the brain is natural or (say) an artificial brain-part replacement that functions equivalently to a natural part. (Mele's useful discussion also recognizes this point, especially on p. 202).

Later in his discussion van Inwagen says:

A change in a natural part of one may well be identical with one's coming to have a certain desire or with one's acquiring a certain belief. But a change in something that is *not* a part of one—even if it is inside one's head and made of brain cells—could not possibly be identical with either of these things. (van Inwagen, 1983, p. 141)

It surely can be granted that a change in something that is *not* a part of one (even if it is inside one) could not be identical with one's having a desire or acquiring a belief. But this is irrelevant, since what is at issue is whether a change in something that is not a *natural* part of one but *is* a part of one that is appropriately integrated into one's functional economy can be identical with one's having a desire or acquiring a belief.

In contrast to van Inwagen's reaction to the slippery slope argument, I would conclude that at the end of the story the agent does not indeed have control over his behavior. The slippery-slope argument seems to me to illustrate in a graphic way the point that if one lacks control when being controlled by someone else, one lacks it when the process leading to one's behavior is structurally similar and yet there is *no* locus of control. But this would not imply that *any* indeterministic process rules out action and control. It seems to me that it is the *particular kind of indeterminism* posited by the story that vitiates control, and not indeterminism *per se*. More specifically, the problem appears to be with the particular way in which the impulses are generated; they are generated in a way which is not appropriately sensitive to the agent's past and the world (not, as I would say, appropriately "reasons-responsive.")

⁸ Robert Kane 1985 and 1989. Recently Kane has written a new book which further develops the themes of his earlier work: Kane 1996 (referred to above in note 3).

⁹ Strawson 1986, esp. pp. 28–29.

¹⁰ Double 1991.

¹¹ Double 1991, pp. 218–221.

¹² I presented the sort of libertarianism in question—with internal, doxastic indeterminacy—and showed how such libertarianism is consistent with the existence of Frankfurt-type examples (in which an agent acts freely and is morally responsible, even though he lacks alternative possibilities), in Fischer 1995.

¹³ Of course, Mele himself believes (unlike the libertarian) that entirely deterministic processes can contain genuine control by the relevant agent. But he is presumably seeking to offer an account of control that is also acceptable to the libertarian.

¹⁴ My own orientation is not to expect knockdown arguments in this area of philosophy. Thus, I am willing to accept certain conclusions on the basis of what I take to be strong arguments that unfortunately fall considerably short of being decisive. Thus, I am a compatibilist about causal determinism and a certain sort of autonomy ("acting autonomously" or "actual-sequence autonomy"), even though I recognize that reasonable persons can legitimately dissent from this view.

References

- Dennett, Daniel C. (1978). *Brainstorms*. Montpelier, Vermont: Bradford Books: 286–289.
- Double, Richard (1991). *The Non-Reality of Free Will*. New York: Oxford University Press.
- Dworkin, Gerald (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Fischer, John Martin (1995). "Libertarianism and Avoidability: A Reply to Widerker," *Faith and Philosophy* 12: 119–125.
- Fischer, John Martin and Ravizza, Mark (1994). "Responsibility and History," *Midwest Studies in Philosophy* 19: 430–451.
- Fischer, John Martin and Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Kane, Robert (1985). *Free Will and Values*. Albany: State University of New York Press.
- Kane, Robert (1989). "Two Kinds of Incompatibilism," *Philosophy and Phenomenological Research* 50: 219–254.
- Kane, Robert (1996). *The Significance of Free Will*. Oxford: Oxford University Press.
- Strawson, Galen (1986). *Freedom and Belief*. Oxford: Clarendon Press.
- van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- Watson, Gary (1977). "Skepticism about Weakness of Will," *Philosophical Review* 86: 316–339.