

©2002 The Institute of Mind and Behavior, Inc.
The Journal of Mind and Behavior
Autumn 2002, Volume 23, Number 4
Pages 413-416
ISSN 0271-0137

Autonomous Agents: From Self-Control to Autonomy. Alfred R. Mele. New York: Oxford University Press, 1995, 271 pages, \$21.95 softcover; and **Liberation from Self: A Theory of Personal Autonomy.** Bernard Berofsky. New York: Cambridge University Press, 1995, xii+270 pages, \$75.00 hardcover.

Reviewed by Christian Perring, Dowling College

Autonomy is a central concept for several different branches of philosophy, notably ethics and political philosophy. While hardly a neglected topic, it remains both controversial and difficult. The two books under consideration have the potential to be important contributions to the literature.

I start with Mele. *Autonomous Agents* is divided into two parts. The first explains the nature of self-control, while the second explains the nature of autonomy. Introductory and linking chapters are devoted to the distinction between self-control and autonomy. Much of the first part (consisting of seven chapters) repeats and expands on discussion that Mele has published elsewhere, both in academic articles and in his books *Irrationality* (Oxford University Press, 1987) and *Springs of Action* (Oxford University Press, 1992). The book would be much more readable if this first part had been condensed down to a few chapters. As it is, the discussion goes over very familiar ground such as *akrasia*, second-order desires, emotion, and action theory. Mele provides a causal account of human action, combining his helpful examples of people trying to control themselves (Sally the smoker trying to quit, Ian who needs to stop watching TV and get back to work on the garden shed, Ed the procrastinating student with a paper to write) with thorough discussion of the philosophical literature (including rebuttals to objections to his past work by other philosophers) and extremely detailed argument for his own view.

The second part holds more interest. Here Mele starts to break new ground (for him). He argues that even an ideally self-controlled person may lack autonomy. His central example to back up this claim involves the unrecognized manipulation of one person's values by another, such as in brainwashing. He goes on to argue for an account of the nature of autonomy that he hopes will be acceptable to both compatibilists and incompatibilists. His method for doing this is to give a compatibilist account and also an incompatibilist account, and then to show that these two accounts converge, except, of course, in their divergent allegiance to compatibilism and incompatibilism.

Requests for reprints should be sent to Christian Perring, Ph.D., Department of Philosophy, Dowling College, Idle Hour Blvd, Oakdale, New York 11769. Email: perringc@dowling.edu

In his six chapters of this second part, Mele addresses several important issues in ethical theory. In the compatibilist part, he discusses the nature of the psychological state of valuing and its relation to desire and belief; the distinction between causation and compulsion; identification and authenticity; the philosophers he tackles include Joel Feinberg, Gerald Dworkin, and Susan Wolf. Most interesting in this part of the book is what he calls an externalist theory of autonomy. He argues, mostly through the use of carefully constructed thought-experiments, that even if a person seems autonomous, we may discover that she is not because her values have been manipulated. That is to say, even if a person is able to scrutinize her own values and is happy with them, these internal psychological abilities do not guarantee autonomy. Just as a burn is not sunburn if the sun did not cause it, so a certain psychological condition does not count as autonomy if it was caused in the wrong sort of way.

Moving on to the incompatibilist side of his disjunction, Mele gives careful discussion of Peter van Inwagen, Robert Kane, and Galen Strawson. For those who, unlike myself, are ready to seriously entertain incompatibilism, Mele gives a thoughtful discussion.

The final chapter of the book is telling. First, Mele considers nonautonomism, the thesis that humans lack autonomy, and uses it as a foil with which to argue for its opposite, the view that we can be autonomous. Finally, Mele draws up a scorecard of advantages and disadvantages for four views: compatibilist belief in autonomy, libertarianism, agnostic autonomism, and nonautonomism. He concludes that agnostic autonomism (which does not commit itself to either compatibilism or incompatibilism, but is committed to the view that we have autonomy) comes out on top.

I have little to say about the strength of Mele's arguments: for the most part they strike me as plausible. But I have two further comments on the book. First, the book strikes me as overly thorough. There were many sections I would have preferred to be assigned to long footnotes or an appendix. It would help greatly for the book to provide a greater sense of priorities: what is new, what is essential to the argument, what can be skipped on first reading, and why Mele's claims are important. Second, I was disappointed that Mele did not do more to explain the consequences of his account of autonomy. His main example concerns imagined and fantastic cases of psychological manipulation. He says nothing about real cases of psychological manipulation in a political context, such as concern critics of totalitarian societies and feminist critics of patriarchy. Neither does he address in any depth the most obvious cases of non-coerced diminished autonomy in real life, people with mental disorders. Showing how his ideas apply to addiction, compulsion, depression, mania and personality disorders would have helped to bring his ideas alive to a much greater degree, and would have shown the power of his theory.

In *Liberation from Self*, Berofsky takes care to attend to the psychological subtleties of autonomy, and explains how he sees this as essential to a satisfactory philosophical account of autonomy. He discusses cases of rigid personality, addiction, weakness of will, and phobias, for example, and compares the approaches of different psychological theories such as psychoanalysis and cognitive-behavioral approaches. While this is all admirable, his view is rather unclear. Obviously, Berofsky's central idea is that autonomy is liberation from self; the difficulty is in spelling out what this amounts to, and how it differs from other theories of autonomy. While many of the details are plain enough, it is difficult to get a sense of how they fit together, or what fundamental idea motivates them.

Berofsky believes that freedom and autonomy are compatible with determinism. Autonomy is a matter of degree. He is clear that full autonomy requires freedom, the capacity for self-assessment, the possession of a value scheme, readiness to evaluate any idea or desire, and capacity for love, friendship, courage, gratitude, trust, and other emotions and character traits that go with the ability to engage with the world. He explains that objectivity is a key component of autonomy; he characterizes this as a "propensity to respond to the world as it is" (p. 203). He is also clear that he rejects the view that autonomy is tied to any particular value scheme or morality, and he highlights the point that an autonomous person should be able to reject morality altogether. Furthermore, he thinks that second-order desires and reflection on one's self have been overemphasized by philosophers. Berofsky explains that what is more important is "the manner in which the agent is immersed in her surroundings" (p. 183). What is elusive in his account is what exact manner of immersion he has in mind.

Readers will find Berofsky's discussion of identification and ownership of mental states most interesting. As already mentioned, he does not embrace Frankfurt's approach. Berofsky suggests, "The judgment that a life is inauthentic is really a value judgment, not a metaphysical one" (p. 99). He says that identification with a desire is endorsement of that desire rather than incorporation of it into one's self. Indeed, he argues that there is no serviceable conception of self (p. 236). The metaphysical issue of which desires are ours is resolved: "The desires which are most truly ours are the ones which play a central role in the mechanisms accounting for our behavior and our mental life" (p. 221). On this view, while an addict may hate and disavow her cravings, those cravings still belong to her, if she acts on them. Lack of autonomy does not consist in being moved by desires that are not one's own. Rather, "The agent lacks autonomy if she would not choose to be an addict were she, as she is now, to make general life choices under optimally cognitive conditions of the sort demanded for rational identification" (p. 223). Thus Berofsky's account of autonomy (or the lack of it) relies on a conditional concerning what the agent would do under certain optimal conditions, rather than on distinguishing between true and false selves or using the apparatus of second-order desires. This seems a rather broad conception of a "manner in which an agent is engaged in her world" (p. 1).

These suggestions of Berofsky are both interesting and important, and should repay study by both philosophers interested in moral psychology and those interested in identity and self. My main complaint is that I wish Berofsky made it easier for his readers to see the flow of his argument and gave more guideposts as to what his most important claims are.

These two books are both significant works containing ideas that deserve further discussion, and they should provoke debate on the nature of personal autonomy and its relation to freedom and determinism, and the right to autonomy in applied ethics and political theory. The style and organization of the writing of both authors may unfortunately have the consequence of making some readers less inclined to read the books in close detail.