Mele, Alfred R. *Autonomous Agents: From Self-Control to Autonomy.* New York: Oxford University Press, 1995. Pp. 271. $49.95 (cloth).

Part 1 of *Autonomous Agents* develops a conception of an ideally self-controlled person and argues that such a person can fall short of personal autonomy. Part 2 addresses what must be added to self-control in order to yield autonomy.

Chapter 1 explains that akrasia is a trait of character exhibited in uncompelled, intentional behavior that goes against the agent's best judgment. The contrary trait, self-control, is exhibited in behavior that conforms to best judgment in the face of temptation. Self-controlled individuals possess both significant motivation to conduct themselves as they judge best and a capacity to do what it takes so to conduct themselves. Self-control may be regional or global, and it comes in degrees. It is important because, though decisive better judgments are formed on the basis of our evaluation of the objects of desire, the motivational force of our desires is not always in accord with our evaluations.

In chapter 2 Mele discusses the causal bearing of decisive better judgments on intentions. Strict akratic action occurs when an agent judges that his doing *A* would be better (morally, aesthetically, etc.) than his doing *B*, yet in the absence of compulsion he intentionally does *B*. Mele argues that there is no nonartificial akrasia-proof species of evaluative reasoning. Evaluative judgments guide conduct, and our best judgments are capable of influencing intention formation. But sometimes, "owing partly to the influence of recalcitrant desires" (p. 29), best judgment does not lead to the formation of a corresponding intention.

Mele next (chap. 3) takes up a problem about the motivation to exercise self-control. If an agent judges it best to *A* but wants more to *B*, where *B*-ing precludes *A*-ing, how can an exercise of self-control be motivationally open to her? Here Mele invokes, among other things, higher-order motivations to bring one's first-order motivations in line with judgment. Consider the smoker who is more motivated to smoke now than not to do so but who judges it better to refrain from smoking. If self-control is helpless against this "preponderant proximal temptation" (p. 42), then the motivation that clashes with better judgment is irresistible; in such cases, self-control can only be exercised in advance. But other than the extreme case, individuals are able to exercise control over desires. Various strategies are discussed, including those designed to minimize the discomfort of unsatisfied appetites (e.g., forgoing cigarettes but using a nicotine patch), redirecting one's attention, and so forth. Extant psychological literature is well utilized here.

Various unorthodox cases are discussed in chapter 4. Mele shows, for example, that weakness of will can trigger behavior that *coincides* with best judgment, and continent action may conflict with best judgment. Consider the young man who has decided to join his friends in robbing a convenience store even though he judges it best not to do so. At the last minute, he loses his nerve. His weakness, fear, prompted behavior that accorded with his best judgment; had he continently overcome the fear, his behavior would have been contrary to best judgment. (Of course, in the latter case he may have exhibited weakness earlier in agreeing to participate in the robbery.) What is common to akratic action against better judgment and akratic action in accord

with better judgment is that a practical commitment is thwarted by noncompelling, competing motivation. Self-control involves a kind of rationality, "internal practical rationality" (p. 80), the coherence of one's intentional behavior with one's own principles and decisive better judgments. Self-control does not ensure that decisions and intentions will support better judgment, as the case of the young man shows; but it increases the likelihood of this.

Chapters 5–7 examine the bearing of self-control on beliefs, emotions, and values and principles, respectively. Akratic believing is motivated believing that violates a doxastic principle that the believer accepts and such believing was avoidable by means of self-control. A plausible doxastic principle is that it is best not to allow what one wants to be the case to determine what one believes to be the case. Mele shows that instances of self-deception might violate this principle and be avoidable. For example, a man continues to believe that his wife is not having an affair in spite of strong evidence to the contrary. Chapter 6 examines analogues of akratic and continent action concerning emotions and feelings. An agent may acquire or continue to have an emotion or feeling that she judges it best not to have (e.g., jealousy). Emotions may be warranted or unwarranted, depending on whether they are appropriate responses to evoking stimuli. Fear in the face of danger is appropriate; fear associated with phobias is not. Control of emotions is not direct, but they are subject to various kinds of indirect control (pp. 106 ff.). Consider Tom, who judges as unwarranted the delight that he experiences as the result of his colleague's professional failures. To rid himself of this, Tom may absorb himself in his own work. Chapter 7 examines the bearing of self-control on agents' values and principles. One can assess one's own values and principles, and it is easy to imagine circumstances that might occasion such an examination (e.g., discovering that a particular moral belief one holds is at odds with the beliefs of several persons whom one greatly admires). Various factors may interfere with an adequate assessment of one's own values, including laziness, aversion to change, and self-interest.

Mele concludes part 1 (pp. 121 ff.) by arguing that even an ideally self-controlled person, an imaginary being, does not have everything that autonomy requires. Autonomy involves at least critical reflection on one's own preferences. Such reflection is guided by values already in place. But if those values are the products of brainwashing or "mind control," then even if the individual is ideally self-controlled he is still not autonomous.

Part 2 asks what must be added to self-control to yield autonomy. Various concepts are defined, including compatibilism, incompatibilism, libertarianism, and determinism. Since compatibilists hold that autonomy is compatible with the truth of determinism and incompatibilists deny this, Mele rightly says that these two will tell a different story about what must be added to self-control to yield autonomy. Chapters 9 and 10 deal with compatibilism, chapters 11 and 12 with libertarianism. Three species of autonomy regarding an agent's pro-attitudes are distinguished: autonomously *developing* a pro-attitude over a period of time; autonomously *possessing* a pro-attitude during a stretch of time; and being autonomous regarding the *influence* of pro-attitudes on one's behavior (p. 138).

Mele begins his discussion of compatibilism by distinguishing between internalist and externalist views of autonomy. On an internalist account, psy-

chological autonomy is wholly an internal matter; externalists say that there is more to being autonomous than what goes on inside a person. One might think that the autonomous person will be able to "shed" his pro-attitudes, where shedding a pro-attitude involves either eradicating it or significantly attenuating it. But this need not be so. For a person may possess a value that is practically unsheddable, yet his commitment to the value may be rational, it need not violate any of his other principles, and it need not lead him to conduct himself against better judgment. External influences on our values are considerable, but *only some* such influences reduce autonomy. One that does is brainwashing. Here (p. 158) Mele appeals to the familiar distinction between causation and compulsion. Suppose that Charles Manson and Beth are "psychological twins." Each holds the same values; but while Manson acquired his values in some usual way, Beth's values were instilled by brainwashers. Beth is the victim of compulsion. These values may be unsheddable for each; yet we would hold Manson responsible, but not Beth. This is because etiology matters; Beth has an "authenticity-blocking history." But internalism cannot capture this; for the difference between Beth and Manson concerns something outside the person.

In chapter 10 Mele develops sufficient conditions for compatibilist autonomy. Factors that can thwart autonomy include compulsion, coercion, being deprived of relevant information, and faulty reasoning skills. Mele argues that what compatibilists will add to an ideally self-controlled agent to yield autonomy are (1) the agent's motivational states are neither compelled* nor coercively produced (where compulsion* is compulsion not arranged by the agent herself), (2) the agent's beliefs are conducive to informed deliberation, and (3) the agent is a reliable deliberator. Satisfaction of conditions 1–3 is compatible with determinism.

Libertarians worry about determinism because they believe its truth implies that our actions are not up to us; instead, they are consequences of the laws of nature and past events. But if there is real autonomy, it must be up to us which of several possible futures comes about. So libertarians hold that there must be indeterministic gaps; if what comes about is up to the agent, "internal indeterminism" is true. But this gives rise to the "control problem" (p. 199), for autonomy and responsibility require that agents have control over their actions and indeterminism seems to weaken "agential" control. In chapter 12 Mele develops a libertarian response to this. What libertarians need is that some relevant events of the central nervous system are causally undetermined; then free choices will not be explicable by external factors. Mele believes that a modest indeterminism will suffice. An illustration is in the doxastic sphere, where all that is undetermined is which members of a shifting subset of an agent's relevant nonoccurrent beliefs will be occurrent and function in his deliberation. Only some beliefs will come to mind, but it is not causally determined which will. During deliberation, it is causally open what the agent will judge best to do. But agents have some control; they are not helpless with regard to the influences that beliefs which do come to mind have on them. Mele is aware that some will think this is too weak, and he tries to answer that (pp. 218 ff.). He concludes that what the libertarian will add to self-control to yield autonomy are the compatibilist's conditions 1–3 plus a fourth: (4) doxastic indeterminism is a regular feature of the agent.

Mele does not choose between compatibilists and libertarians; he simply shows how each can account for autonomy. Nonautonomists hold that no human being is autonomous. In chapter 13 Mele shows how libertarians and compatibilists can combine their resources at least to put the burden of proof on nonautonomists.

Regarding Mele's account of libertarianism, one might wonder if there is an appropriate connection between what is causally undetermined and what is up to the agent. But this is an excellent book. It is rich in arguments, replete with useful examples, and informed by the literature in philosophy and psychology. It is not always easy going, but one's efforts are rewarded. I highly recommend it.

TERRANCE MCCONNELL
University of North Carolina at Greensboro