

Chapter 2

Liberty: Autonomy – Introduction

One plausible way to characterize autonomous agents is to claim that they reflectively endorse their preferences (Dworkin 1988; Frankfurt 1988). But how exactly should we unpack the notion of reflective endorsement? One popular suggestion is Harry Frankfurt's: an agent reflectively endorses a first-order preference when he has a second-order preference that he have that first-order preference (Frankfurt 1988). This suggestion has been met with the objection that one's second-order preferences do not necessarily reflect an authentic self any more so than one's first-order preferences do (Thalberg 1989). In response to this objection, some authors have developed limitations on the history of, and influences on, the formation of first-order preferences that only autonomous agents satisfy (Christman 1991; Mele 1995). But no one has yet succeeded in articulating a precise account of the process through which an autonomous agent determines what preferences to have, and whether to endorse them. In the chapters that follow, I will develop a more robust and more precise characterization of the capacity of autonomy—the capacity to lead an autonomous life, one that is self-ruled or self-governed—than has yet been offered. In these first three chapters, I will be concerned exclusively with the rational dimension of this capacity: the capacity to deliberate over and choose one's own ends, to form intentions to pursue those ends, and to deliberate over and choose means to achieve those ends. My particular concern will be with the first of these—rational deliberation about ends. I address the other dimension of autonomy, self-control, in Chap. 7.¹ I begin here with a brief review of the work of Gerald Dworkin and Joseph Raz on the nature and value of autonomy. An examination of their work provides an excellent starting point for my task, while also bringing the deficiencies of prior accounts of autonomy into focus.

Raz takes autonomy to be “an ideal of self-creation...An autonomous person's well-being consists in the successful pursuits of self-chosen goals and relationships” (Raz 1986, p. 370). In order to possess the capacity of autonomy, a person “must have the mental abilities to form intentions of a sufficiently complex kind,

¹Self-control is the ability to control one's passions and their influence over one's actions, and to stick to one's goals and plans in the face of temptation to do otherwise.

and plan their execution. These include minimal rationality, the ability to comprehend the means required to realize his goals, the mental faculties necessary to plan actions, etc.” (Raz 1986, pp. 372–373). Raz also asserts that “[R]eason affects our choice of ends...just as much as it affects our deliberations” (Raz 1999, p. 73). Raz emphasizes what can be called the competence aspect of autonomy.² The autonomous person, as the author of his own life, must be competent at exercising the capacities for rational thought which Raz enumerates. Raz’s detailed account, however, focuses exclusively on instrumental rationality: the ability of an agent to comprehend, deliberate about, and adopt means to his ends. While acknowledging that reason plays a role in the choice of ends, he does not address the question of how a rational agent should go about determining what ends to adopt.

According to Dworkin, the autonomous person’s ends are self-chosen insofar as his intentions to pursue those goals are formed after critical reflection on his initial preferences, which he then either endorses or changes on the basis of that reflection. On this view, “[A]utonomy is conceived of as a second-order capacity of persons to reflect critically upon their first-order preferences...and the capacity to accept or attempt to change these in light of higher-order preferences and values” (Dworkin 1988, p. 20). Dworkin’s view takes two aspects of autonomy into account. Like Raz, he recognizes the competence aspect of autonomy. Dworkin, however, does have something to say about the role of reason in choosing ends. An autonomous agent is one with the capacity for critical reflection on, and revision of, one’s first-order preferences. This capacity is a prerequisite for the agent’s autonomously forming intentions to pursue ends. Dworkin’s view also recognizes a second aspect of autonomy, which we may call the authenticity aspect. According to Dworkin, an agent has achieved authenticity, in the sense that his preferences are truly his own, once he has brought his first-order preferences into alignment with his higher-order preferences. Authenticity is the result of successfully revising one’s preferences; one achieves it by exercising the competence aspect of autonomy as well as one can. I will have something to say about authenticity at the end of Chap. 4.

What is the exact nature of the competence aspect of autonomy—the capacity to reflect critically on one’s preferences, and then either endorse or revise them on the basis of that reflection? On this point, Dworkin’s account is silent. The guiding thought of the next two chapters is that this capacity is a species of the capacity for practical reasoning. Specifically, it is the capacity for rational deliberation about what ends to adopt, deliberation which is conducted in the light of both the agent’s persisting pre-deliberative attachments, and the evidence he gathers from his experiences of life and the world, which bear on the question of what ends would be best for him to adopt. My ultimate goal for the next two chapters is to develop a precise framework for representing competence in exercising this capacity of

²Raz also discusses another concept—“independence”—which he takes to be an aspect of autonomy. Independence, however, is not an aspect of autonomy, but a type of freedom—freedom from domination, whether physical or psychological, by others. Since I maintain a sharp distinction between autonomy and freedom, I do not discuss this concept in this part of the book. Independence is implied in my concept of freedom to exercise one’s capabilities.

ends-deliberation. This framework will thus represent the core component of the capacity of autonomy. It is ideal competence in exercising this capacity that I will investigate; my framework will characterize the ideal of autonomy, rather than, to borrow Stanley Benn's term, the imperfect capacity of autarchy normally found in actual agents (Benn 1988). As in many other cases, modeling the ideal is the first step toward rigorous understanding, and once that step has been taken a model of the bounded capacity we ordinarily encounter may be possible.

Both Raz's and Dworkin's remarks on the value of autonomy also suggest that the core of the capacity of autonomy is a capacity for ends-deliberation (Dworkin 1988, p. 20; Raz 1986, p. 377). They both locate the value of autonomy in the fact that the autonomous agent gives meaning to his own life by choosing his own goals and commitments and constructing his life-plans for himself. But for these choices to count as the agent's way of constructing the meaning of his life, they must be reasoned choices. An agent cannot be said to determine for himself the meaning of his own life if these choices simply reflect preferences that the agent finds himself having. And it is hard to see how agents could reflectively endorse or revise their preferences in the way required for autonomy, if not as a result of a process of deliberating about what ends they should have for their lives and what options are worth choosing. I suspect that the difficulty of modeling reasoning about ends is responsible for the fact that so many authors with this general view of the nature and value of autonomy have failed to connect the capacity of autonomy with ends-deliberation.

Since the capacity of autonomy is, at least in its rational dimension, the capacity to adopt goals and construct plans to pursue them, it is natural to ask how a theory of autonomy such as mine relates to Michael Bratman's Planning Theory of practical reasoning. Much of the background for my framework for ends-deliberation is shared by the Planning Theory. Bratman describes the structure of plans as partial (insofar as the agent fills in the details of his plan as he proceeds) and as hierarchical (we plan to achieve some goals for the sake of achieving larger ones). This matches my discussion, in Chap. 4, of the nature of an ultimate end, the content of which is filled in as the agent selects the final ends which are constitutive of it. Bratman identifies two reasons for structuring plans in this way: the need to coordinate one's life subject to a limited capacity for reconsideration, and the need to cope with unforeseen events. I accept these as reasons for conceiving of one's ultimate end as an initially thin end that becomes more robust as its constituent ends are selected on the basis of deliberation. I also accept his constraints on plans as constraints on acceptable sets of final ends. These constraints are (1) internal consistency (it should be possible for one to achieve all of one's final ends); (2) strong consistency relative to one's (the beliefs that one's final ends are attainable and are at least as choiceworthy as the other available ends should be consistent with one's other beliefs); and (3) means-ends coherence (as time goes by, the agent should select final ends and acceptable means to those ends).

Bratman's central question is how to assess the rationality of forming an intention. Since intentions are the building blocks of plans, this question has a grander form: how to assess the rationality of the plans an agent makes for his life (Bratman

1987, p. 29). Bratman observes that one characteristic of intentions is that they set ends for further deliberation (Bratman 1987, p. 24). Deliberating about what intention to form, then, can involve deliberating about what end to adopt, not just deliberating about what means to an end to adopt. So Bratman's theory can be taken as an attempt to characterize the same competences in exercising rational capacities as my theory is concerned with—capacities that are integral components of autonomy. Bratman, and I, however, approach this question in very different—though complementary—ways.

The Planning Theory culminates in the Historical Principle of Deliberative Rationality. This principle states that:

If A at t_1 forms the intention to ϕ at t_2 on the basis of deliberation at t_1 , then it is rational of A at t_1 to intend to ϕ at t_2 iff:

- (1) for those intentions of A 's that play a direct role as a background of A 's deliberation, it is rational of A at t_1 so to intend; and
- (2) A reasonably supposes that ϕ is at least as well supported as its relevant, admissible alternatives. (Bratman 1987, p. 85)

My theory of the capacity of autonomy provides specific answers to the questions which this very general principle leaves unanswered. Each part of the principle presumes that it is possible to characterize the variety of practical reasoning that my theory is concerned with. The first part states that the intentions that serve as the background of a new deliberation must themselves be rational. The problem is that Bratman's theory is silent about how to assess the rationality of the agent's background intentions, and of the process of deliberation that leads from rational background intentions to a new rational intention. When the content of an intention is to adopt an end, a theory of ends-deliberation is required to assess the rationality of the intention. The second part of the principle states that the option settled on by the agent must be at least as well supported as the other relevant, admissible options. Again, whenever the agent adopts an end, the assessment of rational support will have to be made in the light of a theory like mine.

References

- Benn, S. 1988. *A theory of freedom*. Cambridge: Cambridge University Press.
- Bratman, M. 1987. *Intention, plans and practical reason*. Cambridge: Cambridge University Press.
- Christman, J. 1991. Liberalism and individual positive freedom. *Ethics* 101(3): 343–359.
- Dworkin, G. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.
- Frankfurt, H. 1988. Freedom of the will and the concept of a person. In *The importance of what we care about*, 11–25. Cambridge: Cambridge University Press.
- Mele, A. 1995. *Autonomous agents: From self-control to autonomy*. New York: Oxford University Press.
- Raz, J. 1986. *The morality of freedom*. Oxford: Clarendon Press.
- Raz, J. 1999. Explaining normativity: On rationality and the justification of reason. In *Engaging reason*, 67–89. Oxford: Oxford University Press.
- Thalberg, I. 1989. Hierarchical analyses of unfree action. In *The inner citadel essays on individual autonomy*, ed. J. Christman, 123–136. New York: Oxford University Press.



<http://www.springer.com/978-3-319-28276-3>

Renewing Liberalism

Sherman, J.A.

2016, XIX, 516 p. 2 illus., Hardcover

ISBN: 978-3-319-28276-3