# *Chapter 10*

# PLANNING AGENCY, AUTONOMOUS AGENCY

## I. PLANNING AND CORE ELEMENTS OF AUTONOMY

Humans seem sometimes to be autonomous, self-governed agents: their actions seem at times to be not merely the upshot of antecedent causes but, rather, under the direction of the agent herself in ways that qualify as

This chapter is to a significant extent an overview of themes I have discussed in a recent series of essays. For further details, see "Identification, Decision, and Treating as a Reason," as reprinted in my *Faces of Intention* (New York: Cambridge University Press, 1999): 185–206; "Reflection, Planning, and Temporally Extended Agency," *Philosophical Review* 109 (2000): 35–61 [this volume, essay 2]; "Valuing and the Will," *Philosophical Perspectives: Action and Freedom* 14 (2000): 249–65 [this volume, essay 3]; "Hierarchy, Circularity, and Double Reduction," in S. Buss and L. Overton, eds., *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge, MA: MIT Press, 2002): 65–85 [this volume, essay 4]; "Nozick on Free Will," in David Schmidtz, ed., *Robert Nozick* (New York: Cambridge University Press, 2002): 155–74 [this volume, essay 6]; "Two Problems about Human Agency," *Proceedings of the Aristotelian Society* 101 (2001): 309–26 [this volume, essay 5]; "Autonomy and Hierarchy," in Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul, eds., *Autonomy* (New York: Cambridge University Press, 2003): 156–76 [this volume, essay 8]; "Shared Valuing and Frameworks for Practical Reasoning," in R. Jay Wallace et al., eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Oxford: Oxford University Press, 2004): 1–27 [this volume, essay 13]; "A Desire of One's Own," *Journal of Philosophy* (2003): 221–42 [this volume, essay 7]; "Three Forms of Agential Commitment: Reply to Cullity and Gerrans," *Proceedings of the Aristotelian Society* 104 (2004): 329–37 [this volume, essay 9]; and "Temptation Revisited," this volume, essay 12. The present essay benefited from written comments from Alfred Mele and Manuel Vargas, and from extremely helpful discussion in a meeting of the Stanford Social Ethics and Normative Theory discussion group and in a colloquium at the University of Miami. It was completed while I was a Fellow at the Center for Advanced Study in Behavioral Sciences. I am grateful for financial support provided by the Andrew W. Mellon Foundation.

a form of governance by that agent. What sense can we make of this apparent phenomenon of governance by the agent herself?[1]

Well, we can take as given for present purposes that human agents have complex psychological economies and that we frequently can explain what they do by appeal to the functioning of these psychological economies. She raised her arm because she wanted to warn her friend; she worked on the chapter because of her plan to finish her book; she helped the stranger because she knew this was the right thing to do; he left the room because he did not want to show his anger. These are all common, everyday instances of explaining action by appeal to psychological functioning. In doing this, we appeal to attitudes of the agent: beliefs, intentions, desires, and so on. The agent herself is part of the story; it is, after all, her attitudes that we cite. These explanations do not, however, simply refer to the agent; they appeal to attitudes that are elements in her psychic economy. The attitudes they cite may include attitudes that are themselves about the agent and her attitudes—desires about desires, perhaps. But what does the explanatory work is, in the end, the functioning of (perhaps in some cases higher-order) attitudes. These explanations are, I will say, nonhomuncular.

When we come to self-governance, however, it is not clear that we can continue in this way. The image of the agent directing and governing is, in the first instance, an image of the agent herself standing back from her attitudes and doing the directing and governing. But if we say that this is, in the end, in what self-governance consists, we will be faced with the question whether the agent who is standing back from these attitudes is herself self-governing. And it is not clear how such an approach can answer that question. Further, if this is, in the end, what we say constitutes self-governance, then it will be puzzling how self-governing human agents can be part of the same natural world as other biological species. Granted, there is already a problem in understanding how the kind of

1. As indicated, I understand self-governance of action to be a distinctive form of self-direction or self-determination (I do not distinguish these last two) of action. Autonomy—that is, personal autonomy—is self-direction that is, in particular, self-governance. Or anyway, that is the phenomenon that is my concern here. (See my "Autonomy and Hierarchy," 156–57, 168 [this volume, pp. 162–63, 177].) Autonomy is related in complex ways to moral responsibility and accountability, but I do not consider these further issues here.

psychological functioning cited in ordinary action explanation can be part of that natural world. But here I assume that we can, in the end, see such explanatory appeals to mind as compatible with seeing ourselves as located in this natural order. But if, in talking of self-governance, we need to see the agent as playing an irreducible role in the explanation of action, we have yet a further problem in reconciling our self-understanding as autonomous with our self-understanding as embedded in a natural order.[2]

These reflections lead to the question of whether there are forms of psychological functioning that can be characterized without seeing the agent herself as playing an irreducible role and that are plausible candidates for sufficient conditions for agential governance. It is also an important question, of course, whether certain forms of functioning are necessary for self-governance. But given the structure of the problem as I have characterized it, the basic issue is one about sufficient conditions for autonomy; and we should be alive to the possibility that there are, at bottom, several different forms of functioning, each of which is sufficient, but no one of which is necessary for self-governance.[3]

In response to this question, the first thing to say is that relevant psychological functioning will involve, but go beyond, purposive agency. Autonomous agents are purposive agents, but they are not simply purposive agents. Many nonhuman animals are purposive agents—they act in ways that are responsive to what they want and their cognitive grasp of how to get it—but are unlikely candidates for self-governance. A model of our autonomy will need to introduce forms of functioning that include but go beyond purposiveness.

In earlier work, I have emphasized that it is an important feature of human agents that they are not only purposive agents; they are also

2. See J. David Velleman, "What Happens When Someone Acts?" in his *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000): 123–43; and R. E. Hobart, "Free Will as Involving Determination and Inconceivable without It," as reprinted in Bernard Berofsky, ed., *Free Will and Determinism* (New York: Harper & Row, 1966): 63–95, esp. 65–66.

3. As for the provision of fully sufficient conditions, though, see my qualifications below in remarks about core elements of autonomy. Alfred R. Mele also pursues a strategy of seeking sufficient (but perhaps not necessary) conditions for certain forms of autonomy. And Mele addresses issues about the historical background of autonomy, issues that, as I explain below, I put aside here. See Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York: Oxford University Press, 1995): 187.

planning agents.[4] Planning agency brings with it further basic capacities and forms of thought and action that are central to our temporally extended and social lives. Indeed, our concept of intention, as it applies to adult human agents, helps track significant contours of these planning capacities. I call my efforts to characterize these features of human agency, and the associated story of intention, the "planning" theory."

As important as it is, however, the step from purposive to planning agency is not by itself a step all the way to self-government. After all, one's planning agency may be tied to the pursuit of ends that are compulsive or obsessive or unreflective or thoughtless or conflicted in ways incompatible with self-government.

This may suggest that though the step from purposive to planning agency is an important step, it is a side step: It does not help us provide relevant sufficient conditions for self-governance. I believe, however, that this suggestion is mistaken, that important kinds of self-governance involve planning attitudes and capacities in a fundamental way.

J. David Velleman once remarked that "an understanding of intention requires an understanding of our freedom or autonomy." And he argued that my 1987 planning theory of intention "falls short in some respects because [it] tries to study intention in isolation from such questions about the fundamental nature of agency."[5] On one natural interpretation of these remarks, the claim is that a theory of intention needs itself to be a theory of autonomy. And this seems too strong to me. There can be intending, planning agents who are not autonomous. A theory of intention should not suppose that only autonomous agents have the basic capacities involved in intending and planning. Nevertheless, I do think that the planning theory of intention has a significant contribution to make to a theory of autonomy.

Let me try to articulate more precisely the kind of contribution I have in mind.[6] We seek models of psychological structures and functioning

4. See my *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987; reissued by CSLI Publications, 1999); and my *Faces of Intention*.

5. See his review of my *Intention, Plans, and Practical Reason* in *Philosophical Review* (1991): 283.

6. See my "Autonomy and Hierarchy," 157 [this volume, pp. 163–64].

that, in appropriate contexts, can constitute central cases of autonomous agency. We should not assume there is a unique such model, but we can consider it progress if we can provide at least one such model. Further, to make progress in this pursuit, we do well, I think, to focus initially on psychological structures and forms of functioning that are more or less current at the time of action, broadly construed. In the end, we will want to know whether there are further constraints to be added, constraints on the larger history of these structures and forms of functioning. Perhaps, for example, certain kinds of prior manipulation or indoctrination need to be excluded. But before we can make progress with that question of history, we need plausible models of important and central structures and functioning on (roughly) the occasion of autonomous action. I will call a model of such important and central structures and functioning a "model of core elements of autonomy." A model of core elements need provide neither necessary nor fully sufficient conditions for autonomy. It need not provide necessary conditions, for it may be that there is more than one way to be autonomous. And it need not provide fully sufficient conditions, for it may be that to ensure autonomy we need also to impose conditions on the larger history. Nevertheless, a plausible model of core elements would help us understand autonomy and its possible place in our natural world.[7] And I want to argue that the planning theory has an important contribution to make to a plausible model of core elements of autonomy.

My argument will take the following form. I will examine two prominent models of relevant forms of psychological functioning: (1) hierarchical models that highlight responsiveness to higher-order conative attitudes; and (2) value-judgment-responsive models that highlight responsiveness to judgments about the good. Although each of these models points to an important form of functioning, each faces problems when offered as a model of core elements of self-governance. My proposal will be that we solve these problems by drawing on the planning theory.

7. And it would be a model of what I have called "core features of human agency." See my "Reflection, Planning, and Temporally Extended Agency," 35–36 [this volume, pp. 21–22]. I point to a similar idea in talking about "strong forms of agency" in "A Desire of One's Own," 222 n. 3 [this volume, p. 138, n. 3].

## 2. THE HIERARCHICAL MODEL AND WATSON'S THREE OBJECTIONS

Let's begin with hierarchy. Here the idea is that the basic step we need to get from mere purposiveness to self-government is the introduction of higher-order conative attitudes about the functioning of first-order motivating attitudes. One main source of this idea is a complex series of papers by Harry Frankfurt.[8] In his classic early essay, Frankfurt wrote that "it is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will."[9] Here, by "will," Frankfurt means, roughly, "desire that motivates action"; and a second-order volition is a second-order desire that a certain desire motivate. When the effective motivation of action (the "will") conforms to and is explained by[10] an uncontested second-order volition, the agent exercises freedom of the will. And when Frankfurt later turns explicitly to autonomy and self-government (which he sees as the same thing), it seems fairly clear that something like this hierarchical story is built into his approach.[11]

Now, we have observed that self-government seems to involve the agent's standing back and doing the governing. The hierarchical model acknowledges the power of this picture, a picture that highlights the agent's reflectiveness about her motivation. But the model goes on to understand such reflectiveness by appeal to certain higher-order attitudes—in the simplest case that Frankfurt initially emphasized, an uncontested

8. See Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988); and *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999). For related ideas, see also Gerald Dworkin, "Acting Freely," *Noûs* 4 (1970): 367–83; Wright Neely, "Freedom and Desire," *Philosophical Review* 83 (1974): 32–54; and Keith Lehrer, "Reason and Autonomy," in Paul, Miller, and Paul, eds., *Autonomy*, 177–98.

9. Frankfurt, "Freedom of the Will and the Concept of a Person," in his *The Importance of What We Care About*, 20. (It is interesting to note that in this passage Frankfurt appeals to something the agent is doing—namely, securing the cited conformity.)

10. Frankfurt points to this condition of explanatory role in his "Identification and Wholeheartedness," in *The Importance of What We Care About*, 163.

11. See esp. Frankfurt's "Autonomy, Necessity and Love" in his *Necessity, Volition, and Love*, 129–41. For a helpful discussion of some issues of Frankfurt interpretation that I am skirting over here, see James Stacey Taylor, "Autonomy, Duress, and Coercion," in Paul, Miller, and Paul, eds., *Autonomy*, 129 n. 5.

second-order volition. In this way, it tries to see self-governance as involving reflectiveness without a homunculus.

Note that the theory need not claim that the very same higher-order attitude is involved in all cases of hierarchical self-governance. It need only claim that all cases of hierarchical self-governance involve some such higher-order conative attitude.

This basic idea has been developed in a number of different ways in recent years both by Frankfurt and by others, and I will later advert to some elements from this literature. But enough has been said about the hierarchical model to see the force of an important trio of objections that were proffered by Gary Watson in response to Frankfurt's initial paper.[12]

Watson's first objection begins with an idea that is central to the hierarchical model, the idea that when a relevant, uncontested higher-order conative attitude favors a certain first-order motivation, the *agent* endorses, or identifies with, that motivation. In the terms of Frankfurt's early version of hierarchy, my uncontested second-order volition in favor of my desire to turn the other cheek constitutes my endorsement of, or identification with, that desire. That is why it is plausible to say that when that desire motivates action, in part because of my second-order volition, *I* am directing my action. But, Watson observes, the hierarchical model does not seem to have the resources to explain this. After all,

> since second-order volitions are themselves simply desires, to add them to the context of conflict is just to increase the number of contenders; it is not to give a special place to any of those in contention.[13]

We can express the point by saying that there is nothing in the very idea of a higher-order desire that explains why it has authority to speak for the agent, to constitute where the agent stands. For all that has been said, when action and will conforms to a higher-order desire, it is simply

12. Gary Watson, "Free Agency," *Journal of Philosophy* 72 (1975): 205–20. R. Jay Wallace endorses similar objections in his "Caring, Reflexivity, and the Structure of Volition," in Monika Betzler and Barbara Guckes, eds., *Autonomes Handeln* (Berlin: Akademie Verlag, 2000): 218–22.

13. Watson, "Free Agency," 218.

conforming to one attitude among many of the wiggles in the psychic stew. The hierarchical model does not yet have an account of the *agential authority* of certain higher-order attitudes.[14] But it needs such an account in order to provide a nonhomuncular model of agential governance. And that is Watson's first objection.[15]

Watson's second objection is built into the alternative model he offers, a model that highlights responsiveness to judgments of the good. Watson sees such judgments as an "evaluational system" that "may be said to constitute one's standpoint."[16] If we are looking for attitudes that speak for the agent, that constitute where the agent stands, then the natural candidates are not higher-order volitions, but evaluative judgments about what "is most worth pursuing."[17] I will call this idea, that the agent's standpoint is constituted by evaluative judgment rather than by higher-order conative attitude, the "Platonic challenge" to the hierarchical model.

Watson's third objection draws on but goes beyond this. He writes:

> [Agents] do not (or need not usually) ask themselves which of their desires they want to be effective in action; they ask themselves which course of action is most worth pursuing. The initial practical question is about courses of action and not about themselves.[18]

Here Watson is emphasizing his Platonic model; but he is also pointing to a further objection, one that involves a claim about the structure of ordinary deliberation. The basic idea is that ordinary deliberation is first-order

14. Talk of agential authority comes from my "Two Problems about Human Agency"; talk of wiggles in the psychic stew comes, I admit, from my "Reflection, Planning, and Temporally Extended Agency," 38 [this volume, p. 24].

15. Watson notes that there are elements in Frankfurt's essay—in particular, Frankfurt's talk of an agent who "identifies himself *decisively* with one of his first-order desires"—that suggest that it is not conative hierarchy that is doing the main theoretical work but, rather, the idea of decisive identification. But, Watson remarks, if "notions of acts of identification and of decisive commitment . . . are the crucial notions, it is unclear why these acts of identification cannot themselves be of the first order." (The quote from Frankfurt is in Watson's "Free Agency," at 218, while the quote from Watson is at 219.) I discuss this exchange between Frankfurt and Watson in "Identification, Decision, and Treating as a Reason," in my *Faces of Intention*, 188–90.

16. Watson, "Free Agency," 216.

17. Ibid., 219.

18. Ibid.

deliberation about what to do, not higher-order reflection about one's desires. And the objection is that the hierarchical model misses this point and mistakenly sees deliberation as primarily a matter of higher-order reflection on motivating attitudes. Let us call this the "objection from deliberative structure."

So we have a trio of objections to the hierarchical model: the objection about agential authority, the Platonic challenge, and the objection from deliberative structure. Taken together, these constitute a serious challenge to the hierarchical model.

## 3. THE PLATONIC MODEL AND UNDERDETERMINATION BY VALUE JUDGMENT

I want to give the hierarchical model something to say in response to this challenge. My strategy is to do this by bringing together elements from the hierarchical model with elements from the planning theory. Before proceeding with this strategy, however, I want to reflect on the Platonic alternative that Watson sketches, one that highlights responsiveness to judgments about the good.

An initial observation is that it seems possible for one to judge that, say, turning the other cheek is best, but still be alienated from that judgment in a way that undermines its agential authority.[19]

We can clarify one way this can happen by turning to one of Frankfurt's later developments of the hierarchical model. In response to concerns about what I have called "agential authority," Frankfurt introduced an important idea: satisfaction.[20] Satisfaction is not a further attitude, but rather a structural feature of the psychic system. For me to be satisfied with my higher-order desire in favor of my desire to turn the other cheek is not for me to have an even-higher-order desire. It is, rather, for my higher-order desire to be embedded in a psychic system in which there is no relevant tendency to change: "Satisfaction is a state of the entire psychic system—a state constituted just by the absence of any tendency or

19. Frankfurt made this point in conversation. Also see Velleman, "What Happens When Someone Acts?" 134.

20. Frankfurt, "The Faintest Passion," in his *Necessity, Volition, and Love*, 103–5.

inclination to alter its condition."[21] Frankfurt's idea—expressed in the terms I have introduced here—is that such a higher-order desire has agential authority when the agent is satisfied with it.

I have elsewhere noted that satisfaction with such a desire may be grounded in depression, and in such cases satisfaction with desire does not seem to be enough to guarantee agential authority.[22] Nevertheless, I think that this idea of satisfaction is important in two ways. First, a version of it will be of use later, as one part of a more adequate account of agential authority. Second, it helps us see that one may be dissatisfied with, and for that reason alienated from, one's evaluative judgment in a way that undermines its agential authority. This is one way in which the Platonic proposal is faced with a problem of agential authority.

However, a defender of the Platonic proposal can, in response, focus on evaluative judgments with which the agent is, in an appropriate sense, satisfied. She may then propose that it is such evaluative judgments that constitute the agent's standpoint. A full defense of this proposal would need to say more about the roles of such evaluative judgments in our agency and why these help establish agential authority. Nevertheless, this does show how the Platonic model can, like the hierarchical model, draw on the idea of satisfaction.

But now we need to consider a different kind of alienation from value judgment, one that was emphasized by Watson himself in a later essay.[23] One might have a settled judgment that turning the other cheek would be best, might be satisfied with that as one's settled evaluative judgment, but nevertheless be fully committed, rather, to revenge. As Watson says, "I might fully 'embrace' a course of action I do not judge best." Watson calls such situations "perverse cases." In such cases, the agent's "standpoint" is not captured by his evaluative judgment but rather by his "perverse" commitment.

However, while Watson was right to emphasize such cases, a defender of the Platonic model does have a response. She can say that such cases

21. Frankfurt, "The Faintest Passion," 104.

22. Bratman, "Identification, Decision, and Treating as a Reason," 194–95. And see Bratman, "Reflection, Planning, and Temporally Extended Agency," 49 [this volume, p. 35], for my strategy for avoiding this difficulty within my own account.

23. Watson, "Free Action and Free Will," *Mind* 96 (1987): 150. Also see my "A Desire of One's Own," 227 [this volume, p. 144].

involve a rational breakdown and that in the absence of rational breakdown an agent's standpoint consists of relevant evaluative judgments. Because we are seeking conditions for self-government and because the kind of rational breakdown at issue can plausibly be seen as blocking self-governance, this proposal keeps open the idea that self-governance consists primarily of rational responsiveness to relevant evaluative judgments.

This takes me to a third concern—namely, that even in the absence of rational breakdown, the agent's evaluative judgments frequently underdetermine important commitments. Faced with difficult issues about what to give weight or significance to in one's life, one is frequently faced with multiple, conflicting goods: Turning the other cheek is a good, but so is an apt reactive response to wrongful treatment; resisting the use of violence by the military is good, but so is loyalty to one's country; human sexuality is a good, but so are certain religious lives of abstinence. In many such cases, the agent's standpoint involves forms of commitment—to draft resistance, say—that have agential authority but go beyond his prior evaluative judgment. This may be because the agent thinks that, though he needs to settle on a coherent stance, the conflicting goods are more or less equal. Or perhaps he thinks he simply does not know which is more important. (He is, after all, like all of us, a person with significant limits in his abilities to arrive at such judgments with any justified confidence.) Or perhaps he thinks that the relevant goods are in an important way incommensurable.[24] In such cases, there need not be a rational breakdown but rather a sensible and determinative response to ways in which one's value judgments can underdetermine the "shape" of one's life.[25] One may be committed to building into the fabric of one's own life some things one judges good, but not others. And even in a case in which one judges that, say, a life of helping others is strictly better than a life in which one does not help others, one's judgment will typically leave in its wake significant

24. For this last point, see Joseph Raz, "Incommensurability and Agency," as reprinted in his *Engaging Reason* (Oxford: Oxford University Press, 1999): 46–66. I discuss this trio of possibilities in "A Desire of One's Own."

25. See, for example, Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981): 446–50. Talk of the shape of a life comes from Charles Taylor, "Leading a Life," in Ruth Chang, ed., *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press, 1997): 183.

underdetermination of the exact extent to which this value is to shape one's life, the exact significance this value is to have in one's deliberations.

In these cases of underdetermination by prior value judgment, the hierarchical model seems to be in a better position than the Platonic model. The hierarchical model has room for the view that these elements of the agent's standpoint—elements of commitment in the face of underdetermination by prior value judgment—are constituted by relevant higher-order conative attitudes.[26] Granted, we are still without a full account of the agential authority of those higher-order attitudes. But that is not a defense of the Platonic model. Rather, it is an observation that, so far, neither model solves the problem of agential authority.

It is here that we do well to turn to the planning theory.

## 4. PLANNING, TEMPORALLY EXTENDED AGENCY, AND AGENTIAL AUTHORITY

A basic feature of adult human agents is that they pursue complex forms of cross-temporal and social organization and coordination by way of planning. They settle on—commit themselves to—prior and typically partial and hierarchically structured[27] plans of action, and this normally shapes later practical reasoning and action in ways that support cross-temporal organization, both individual and social. Such plan-like commitments can involve settling matters left indeterminate by prior evaluative judgment, as when one decides on one of several options, no one of which one sees as clearly superior. Indeed, one can be settled on certain intentions, plans, or policies without reflecting at all on whether they are for the best or making an explicit decision in their favor.[28]

According to the planning theory, our planning agency brings with it distinctive norms of plan consistency, plan coherence, and plan stability.

26. For a somewhat similar view, see Keith Lehrer, *Self Trust: A Study of Reason, Knowledge, and Autonomy* (Oxford: Oxford University Press, 1997): chap. 4.

27. The hierarchies I allude to here are, roughly, ones of ends and means, not the conative hierarchies on which I have so far been focusing.

28. In a version of this sort of case emphasized by Nadeem Hussain, an agent in a strongly traditional society unreflectively internalizes certain general policies passed down by the tradition.

To intend to do something in the future or to have a policy concerning certain recurring types of circumstances is to have an attitude that is to be understood in terms of such planning capacities and norms. Such intendings and policies are importantly different from ordinary desires. But they are no more mysterious than the familiar phenomena and norms involved in planning. In this way, the planning theory is a modest, nonmysterious theory of the will.[29]

An agent's plan-like attitudes support cross-temporal organization of her practical thought and action, and they do this in a distinctive way. Prior plans involve reference to later ways of acting; and in filling in and/or executing prior plans one normally sees oneself in ways that refer back to those prior plans. Such plans are, further, typically stable over time. So planning agency supports cross-temporal organization of practical thought and action in the agent's life in part by way of cross-temporal referential connections and in part by way of continuities of stable plans over time. So it supports such organization in part by way of continuities and connections of a sort that are highlighted by Lockean accounts of personal identity over time.[30] And this is no accident: It is a characteristic feature of the functioning of planning in our temporally extended lives.

This opens up an approach to agential authority. The problem of agential authority is the problem of explaining why certain attitudes have authority to constitute the agent's practical standpoint. So far, we have been thinking of this as a problem about the agent at a particular time. But the human agents for whom this problem arises are ones whose agency extends over time: They begin overlapping, and interwoven plans and projects, follow through with them, and (sometimes) complete them. Such temporal extension of agency involves activities at different times performed by the very same agent. A broadly Lockean story of that sameness of agency over time will emphasize relevant psychological connections and continuities. In particular, our planning agency constitutes and supports the cross-temporal organization of this temporally extended agency by way of Lockean connections and continuities—by way of Lockean ties. And this gives relevant

29. See my "Introduction," *Faces of Intention*, 5.

30. See Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984): 206–8; and my "Reflection, Planning, and Temporally Extended Agency," 43–45 [this volume, pp. 28–30].

plan-type attitudes a claim to speak for the temporally persisting agent. As I once wrote, the idea is that "we tackle the problem of where the agent stands *at a time* by appeal to roles of attitudes in creating broadly Lockean conditions of identity of the agent *over time*."[31] And central among the relevant attitudes are plan-type attitudes.

If this is right, then it is good news for the hierarchical theorist. She can see the relevant higher-order conative attitudes—those that constitute the agent's practical standpoint—not merely as desires but rather as plan-type attitudes. She can then cite the Lockean roles of these plan-type attitudes to explain their agential authority. Or, at least, this will be the basic step in such an explanation. In this way, the planning theory can give the hierarchical theorist something more to say in response to the objection from agential authority. And given that intentions and plans are sometimes formed in the face of underdetermination by prior value judgment, such plan-type attitudes are natural candidates to respond to the issues raised by such cases of underdetermination.

## 5. SELF-GOVERNING POLICIES

But what plan-type attitudes are these? Given the role they need to play within the theory we are developing, they need to be higher-order plan-like attitudes. And they need to be higher-order plan-like attitudes that speak for the agent because they help constitute and support the temporal extension of her agency. They will do this in large part by being plan-type attitudes whose primary role includes the organization of practical thought and action over time by way of Lockean ties. This makes it plausible that in the clearest cases the relevant attitudes will be policy-like: They will concern, in a more or less general way, the functioning of relevant conative attitudes over time, in relevant circumstances.[32]

31. Bratman, "Reflection, Planning, and Temporally Extended Agency," 46 [this volume, p. 32].

32. Granted, there will be cases in which a relevant intention-like attitude will be a "singular commitment" to treat a certain desire in a relevant way on *this* occasion. (See my "Hierarchy, Circularity, and Double Reduction," 78–79 [this volume, pp. 68–88].) Such intention-like attitudes will have some claim to agential authority. Given the singularity of the commitment, however, these intention-like attitudes will have a less extensive tie to temporally extended agency and thus a lesser claim to authority. Because our concern is primarily with sufficient conditions for autonomy, I will here put such cases to one side.

What the hierarchical theorist will primarily want to appeal to, then, are higher-order policy-like attitudes. Which higher-order policy-like attitudes? Here we need to reflect further on the very idea of self-governance.

Autonomous actions, I have said, are under the direction of the agent in ways that qualify as a form of governance by that agent. But what forms of agential direction constitute agential governance? Well, the very idea of governance brings with it, I think, the idea of direction by appeal to considerations treated as in some way legitimizing or justifying. This contrasts with a kind of agential direction or determination that does not involve normative content. And this means that the higher-order policy-like attitudes that are cited by the hierarchical theorist should in some way reflect this distinctive feature of self-governance.

Recall Frankfurt's notion of a second-order volition: a desire that a certain desire motivate. The content of such a second-order volition concerns a process of motivation, not—at least not directly—a process of reasoning that appeals to legitimizing, justifying considerations. So such a higher-order attitude does not seem to reflect the way in which self-governance is a kind of governance, not a kind of direction that involves no normative content.

Consider now a higher-order policy concerning a desire for X. One such policy will say that this desire is to influence action by way of practical reasoning in which *X*, and/or the desire for *X*, is given justifying weight or significance. Call such a higher-order policy—one that favors such functioning of the desire in relevant motivationally effective practical reasoning—a *self-governing policy*. Our reflections about self-governance—in contrast with nonnormative self-direction—suggest that self-governing policies can play a basic role in hierarchical theories of self-governance.[33] For reasons we have discussed, such policies have a presumptive claim to agential authority, to speaking for the temporally persisting agent. And such policies will concern which desires are to be treated as providing justifying considerations in motivationally effective practical reasoning. They will in that

33. There will also be room for attitudes that play the higher-order policy-like roles in one's temporally extended agency that I have been emphasizing, though they are not general intentions. I call these "quasi-policies." See my "Reflection, Planning, and Temporally Extended Agency," 57–60 [this volume, pp. 42–44].

sense say which desires are to have for the agent what we can call "subjective normative authority"; and they will constitute a form of valuing that is different from, though normally related to, judging valuable.[34]

Can the hierarchical theory, then, simply appeal to such self-governing policies in its model of self-governance? Well, if the guidance by these policies is to constitute the agent's governance, then we should require that the agent knows about this guidance.[35] Does that suffice? Not quite. Although such policies have a presumptive claim to agential authority, it still seems possible to be estranged from a particular self-governing policy. This is a familiar problem for a hierarchical theory. But we have already noted a further resource available to such a theory: a version of the Frankfurtian idea of satisfaction. To have agential authority, we can say, a self-governing policy must be one with which the agent is, in an appropriate sense, satisfied.[36]

But what if the satisfaction is grounded in depression? Depression might substantially undermine the normal functioning of these self-governing policies. Such a case would not challenge the present account. But what if these self-governing policies continue to play their characteristic roles in Lockean cross-temporal organization—by way of shaping temporally extended deliberation and action—but the absence of pressure for change in those policies is due to depression? Well, in this case, the self-governing policies remain settled structures that play these central Lockean roles in

34. For the point about valuing, see my "Valuing and the Will" and "Autonomy and Hierarchy." For the idea of subjective normative authority, see my "Two Problems about Human Agency." (In section 7, I will be extending this notion of subjective normative authority.) Note that these policies concern the agent's practical *reasoning*. So we need to understand the reasoning that is the focus of these policies in a way that does not reintroduce worries about a homunculus. See my "Hierarchy, Circularity, and Double Reduction," 70–78 [this volume, pp. 74–85]; and "Two Problems about Human Agency," 322–23 [this volume, pp. 90–92].

35. See Garrett Cullity and Philip Gerrans, "Agency and Policy," *Proceedings of the Aristotelian Society* 104 (2004): 317–27, and my "Three Forms of Agential Commitment: Reply to Cullity and Gerrans." This self-knowledge requirement is doubly motivated, by the way. It is a straightforwardly plausible condition on self-governance that the agent know what higher-order policies are guiding her thought and action. But, as Agnieszka Jaworska has noted, it is also unlikely that an unknown policy will have the kinds of referential connections to prior intentions and later action that are central to our Lockean account of agential authority.

36. My efforts to spell out an appropriate sense appear in my "Reflection, Planning, and Temporally Extended Agency," 49–50, 59–60 [this volume, pp. 35–36, 44].

temporally extended, deliberative agency, and they do that in the absence of relevant pressure for change. So it seems to me that they still have a presumptive claim to establish the (depressed) agent's standpoint.

Can we stop here? Can we say that in a basic case self-governance consists primarily in the known guidance of practical thought and action by self-governing policies with which the agent is satisfied? Well, there does remain a further worry: Does self-governance require not just that the agent know about this functioning of the self-governing policy and be satisfied with it, but, further, that the agent *endorse* it in a way that is not just a matter of being satisfied with it? But what could such further endorsement be? Some yet further, distinct, and yet-higher-order attitude? But that way lies a familiar regress.

I think that a natural move for the hierarchical theorist to make at this point is to appeal to reflexivity: The self-governing policies that are central to the model of autonomy that we are constructing will be in part about their own functioning.[37] Such a policy will favor treating certain desires as reason-providing as a matter of this very policy.[38] The idea is not that such reflexivity by itself establishes the agential authority of the policy. Agential authority of such attitudes is, rather, primarily a matter of Lockean role and satisfaction. But in a context in which these conditions of authority are present, a further condition of reflexivity ensures, without vicious regress, the endorsement of self-governing policy that seems an element in full-blown self-governance.

The proposed model, then, appeals to practical reasoning and action that are appropriately guided by known, reflexive, higher-order self-governing policies with which the agent is satisfied. By combining the resources of the hierarchical and the planning theories in this way, we arrive at a nonhomuncular model of core elements of autonomy.

37. I think there is also another reason for such reflexivity, one associated with the concern about reasoning to which I allude in note 34 and the essays cited there.

38. A closely related idea is in Keith Lehrer, *Self-Trust,* 100–102; and also in his "Reason and Autonomy," 187–91. For the basic idea of seeing intentions as reflexive, see Gilbert Harman, *Change in View* (Cambridge, MA: MIT Press, 1986): 85–88. However, my appeal here to reflexivity is not part of a view that, like Harman's, sees *all* "positive" intentions in this way. Further, because my appeal to reflexivity is against a background of a Lockean story of agential authority, together with a Frankfurtian appeal to satisfaction, the job of such reflexivity within my account of autonomy is considerably more limited than its job within Lehrer's.

## 6. REPLIES TO WATSON'S THREE OBJECTIONS

How does this proposed model respond to the cited trio of objections to the hierarchical theory? Well, the response to the objection from agential authority has already been front and center. Higher-order self-governing policies have an initial claim to speak for the temporally persisting agent given their systematic role in constituting and supporting the cross-temporal organization of practical thought and action by way of Lockean ties. This claim is relevantly authoritative when the agent is satisfied with these policies and they have the cited reflexive structure.

What about the Platonic challenge? Here the answer is that we need to be able to appeal to a central and important kind of commitment that goes beyond prior value judgment, given phenomena of underdetermination of the shape of one's life by such judgments. We need to be able to appeal to commitments in the face of judgments of roughly equal desirability or of incommensurability; and we need to be able to appeal to commitments in the face of reasonable inability to reach, with confidence, a sufficiently determinative judgment of value. Indeed, such commitments may arise even in an agent who does not much go in for value judgment. The appeal to self-governing policies provides for such commitments—commitments that will normally have a kind of stability over time that is characteristic of such attitudes.[39]

One way to see what is going on here is to suppose, with a wide range of philosophers, that evaluative judgments are in some important sense subject to intersubjectivity constraints. In contrast, the commitments that constitute an agent's own standpoint need not be subject to such constraints.[40] In cases of underdetermination by value judgment, the agent may sensibly arrive at further commitments that he does not see as intersubjectively directed or accountable in ways characteristic of value

39. I should emphasize that the relevant notion of stability here is in part a normative one: it will involve norms of reasonable stability. It is an important question how exactly to understand such reasonable stability. For some efforts in this direction, see my "Toxin, Temptation, and the Stability of Intention," in my *Faces of Intention* and "Temptation Revisited," this volume, essay 12. Note that the appeal to reasonable stability is *not* an appeal to "volitional necessities" in the sense invoked by Frankfurt in his "Autonomy, Necessity, and Love," 138.

40. For references and further discussion, see my "A Desire of One's Own."

judgment. This leaves open the idea that self-governance precludes a severe breakdown between evaluative judgments with which the agent is satisfied and the commitments that constitute the agent's standpoint. Such a breakdown—as in a Watsonian "perverse" case—is a significant kind of internal incoherence. So it is plausible to say that there is not the kind of unity of view that is needed for self-governance. Nevertheless, and contrary to the Platonic challenge, a model that appeals only to evaluative judgment does not yet provide the resources to characterize forms of agential commitment that are central to self-governance.

What about the objection from deliberative structure? Should our hierarchical model reject Watson's suggestion that "the initial practical question is about courses of action"? Well, sometimes in deliberation one does reflect directly on one's motivation. Nevertheless, I think that Watson is right that frequently in deliberation what we explicitly consider is, rather, what to do. But this need not be an objection to our hierarchical model. We can understand that model as one of background structures that bear on an agent's efforts to answer this "initial practical question": when a self-governing agent grapples with this question, her thought and action are structured in part by higher-order self-governing policies.[41] Or, at least, this is one important case of self-governance.

Those, anyway, are the basic responses to the three objections. But these responses do point to a further issue. We have seen why appeal to higher-order conative attitudes need not be incompatible with the typically first-order structure of ordinary deliberation. We have seen how to explain why certain kinds of higher-order conative attitudes can have agential authority. And we have seen reason for a model of central cases of self-governance to include forms of commitment, to modes of practical reasoning and action,

41. In seeing deliberation as primarily first-order, but also seeing the valuings that enter into deliberation as involving conative hierarchy, my view is in the spirit of certain aspects of Simon Blackburn's approach to these matters. (I provide a different treatment of the relevant hierarchy, however. And my view remains neutral with respect to the basic debate between cognitivist approaches and expressivist approaches of the sort championed by Blackburn.) See Blackburn, *Ruling Passions: A Theory of Practical Reasoning* (Oxford and New York: Clarendon/Oxford University Press, 1998). (Blackburn's remarks about a "staircase of practical and emotional ascent" are at 9; his remarks about valuing are at 67–68; and his remarks about deliberation are at 250–56.)

that go beyond evaluative judgment. But none of these points as yet fully explains the basic philosophical pressure for the introduction of hierarchy into the model. They do show that once hierarchy is introduced, we can respond to challenges concerning agential authority and the structure of deliberation. And they do show that appeal to hierarchical conative attitudes is one way to resolve issues raised by underdetermination by value judgment. But they do not yet fully clarify why we should appeal to such hierarchical attitudes in the first place. Perhaps, instead, we should appeal only to certain first-order plan-like commitments that resolve the problems raised by underdetermination by value judgment, guide first-order deliberation, and also allow for a story of agential authority.

We might respond by reminding ourselves that our fundamental concern is with nonhomuncular sufficient conditions for self-governance. So we need not claim that hierarchy is necessary for self-governance. And this response is correct as far as it goes. But even after noting the availability of this response, there is an aspect of the objection to which we need to respond directly. We need to explain why we should see conative hierarchy as even one among perhaps several different models of core elements of autonomy; and to do that, we need to say more about the pressures for introducing such hierarchy.

This is a salient issue in part because it may seem that the account of self-governance as so far developed lends itself to a modification that leaves the account pretty much intact, but in which conative hierarchy drops out.[42] The idea here would be to appeal to policies simply to give weight or significance to consideration X in one's motivationally effective practical reasoning. Such policies seem to be first-order: Their target is a certain activity of reasoning. But in other respects, it seems they could have the features of self-governing policies that have been exploited by the model: Lockean role in cross-temporal organization, targets of self-knowledge and satisfaction, agential authority, and commitments concerning subjective normative authority that do not require determination by value judgment. So we may wonder why hierarchy should be built into the account. Why not throw away the ladder?

42. As Samuel Scheffler and others have noted in correspondence and conversation.

## 7. REASONS FOR HIERARCHY

We can begin by recalling one reason we have already seen for introducing a kind of conative hierarchy into a model of autonomy: relevant policies about practical reasoning will reflexively support themselves. This is a kind of conative hierarchy. But it is only a limited form of hierarchy, one that does not yet include the idea that such policies are generally about further, distinct forms of first-order motivation. In contrast, hierarchical theories of the sort we have been discussing involve these broader hierarchies of conative attitudes about conative attitudes.[43] So we are still faced with the question of why we should see such broader hierarchies as central to our model of core elements of autonomy.[44]

In at least one strand of his work, Frankfurt's appeal to conative hierarchy is driven by what he takes to be a reflective agent's project of self-constitution. Frankfurt seeks a notion of "internal" that fits with Aristotle's idea that "behavior is voluntary only when its moving principle is inside the agent." And Frankfurt's idea is that "what counts . . . is whether or not the agent has constituted himself to include" a certain "moving principle."[45] The reflective agent's effort at self-constitution is a response to the question, "with respect to each desire, whether to identify himself with it or whether to reject it as an outlaw and hence not a legitimate candidate for satisfaction."[46] In this way, conative hierarchy is seen as involved in the kind of self-constituted internality that is basic to reflective agency.

43. Gilbert Harman argues that (1) "positive intentions are self-referential," so (2) all creatures who have positive intentions have higher-order conative attitudes, and so (3) "Frankfurt's appeal to second-order volitions is not the key to distinguishing autonomy from nonautonomy." Though I would not defend a simple appeal to second-order volitions as this "key," my remarks in the text do point to a response on Frankfurt's behalf to this criticism. Frankfurt can say that what provides the key is the capacity for *broad* conative hierarchy, a capacity that goes beyond the hierarchy built into the purported reflexivity of positive intentions. See Gilbert Harman, "Desired Desires," as reprinted in his *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2000): 122–26.

44. For ease of exposition, in the discussion to follow of reasons for broad hierarchy, I will simply speak of hierarchy where I mean broad hierarchy. Also, I do not claim that the pressures to be discussed exhaust the field. There may be other pressures for conative hierarchy that would need to be considered in a more extensive discussion.

45. Frankfurt, "Identification and Wholeheartedness," 171.

46. Frankfurt, "Reply to Michael E. Bratman," in Sarah Buss and Lee Overton, eds., *Contours of Agency* (Cambridge, MA: MIT Press, 2002): 88.

A second pressure in the direction of conative hierarchy comes from a picture of deliberation as reflection on one's desires, reflection aimed at choosing on which desire to act.[47] Such a model of deliberation, coupled with a search for a nonhomuncular story, can lead straightway to conative hierarchy.

Granted, these two different pressures can interact. Given such a model of deliberation, one may be led to think of deliberation as concerned with self-constitution. And given a Frankfurtian, hierarchical story of self-constitution, one may want to extend it to a model of deliberation.[48] Nevertheless, it is useful to keep these two ideas apart.

One reason this is useful is that these different approaches interact differently with Watson's objection to a model of deliberation as higher-order reflection. Here my strategy has been to argue that—though some deliberation does have this higher-order structure—the hierarchical model of self-governance need not see this as the central case of deliberation. Does this mean that our basic reason for building hierarchy into our model of self-governance should be a metaphysical concern with internality and self-constitution?

Although the issues are complex, I believe that if we stop here we may miss an important practical pressure in the direction of conative hierarchy.

An initial point—from Agnieszka Jaworska—is that the Lockean model of agential authority points to an account of internality (in the sense relevant to the cited Aristotelian idea) that does not make hierarchy essential.[49] There can be important attitudes—a young child's love for her father, say—that do not involve conative hierarchy but nevertheless play the kind of Lockean roles in cross-temporal organization of thought and action that establish

47. Though Christine Korsgaard shares with Frankfurt an interest in self-constitution, she also embraces such a model of deliberation when she writes: "When you deliberate, it is as if there were something over and above all your desires, something which is *you*, and which *chooses* which desire to act on" (*The Sources of Normativity*, 100). For Korsgaard's concerns with self-constitution, see her "Self-Constitution in the Ethics of Plato and Kant," *Journal of Ethics* 3 (1999): 1–29.

48. Though Frankfurt himself does not seem so inclined. (See his "Reply to Michael E. Bratman," 89–90.)

49. See her "Caring and Internality," *Philosophy and Phenomenological Research* (forthcoming). The example to follow comes (with a change in gender) from that paper.

internality. So the concern with internality does not, on its own, provide sufficient philosophical pressure for conative hierarchy.

A Frankfurtian response would grant the point but insist that, *for agents who are sufficiently reflective to be self-governing*, internality of first-order motivation is (normally?) the product of higher-order reflection and higher-order endorsement or acceptance. And this brings with it conative hierarchy. So, while conative hierarchy need not be involved in all cases of internality, internality within the psychology of reflective self-governance needs conative hierarchy.

But now consider an alternative model of reflectiveness. This model highlights first-order policies about what to treat as a reason in one's motivationally effective practical reasoning; and it says that such policies are reflectively held when they are appropriately tied to (even if underdetermined by) evaluative reflection. Here we have a central role for plan-type commitments concerning practical reasoning (to which we can extend our account of agential authority); and we have a kind of reflectiveness; but we do not yet have conative hierarchy.

What this alternative model fails fully to recognize, however, is that human agents have a wide range of first-order motivating attitudes in addition to such first-order policies about practical reasoning and that these other motivating attitudes threaten to undermine these policies. The point is related to an aspect of Aristotle's moral psychology that has been highlighted by John Cooper. Cooper emphasizes that a central Aristotelian theme is that human agents are subject to significant motivational pressures that do not arise from reflection on what is worth pursuing.[50] For our purposes here, what is important is the related idea that human agents are subject to a wide range of motivational pressures that do not arise primarily from their basic practical commitments. Indeed, as we all learn, these motivational pressures may well be contrary to those commitments. The clearest cases include (but are not limited to)

50. John Cooper, "Some Remarks on Aristotle's Moral Psychology," reprinted in his *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory* (Princeton, NJ: Princeton University Press, 1999): 237–52. As Cooper puts the view, "non-rational desires will be desires no part of the causal history of which is ever any process (self-conscious or not) of investigation into the

certain bodily appetites and certain forms of anger, rage, humiliation, indignation, jealousy, resentment, and grief. It is an important fact about human agents—one reflected in our commonsense self-understanding—that such motivating attitudes are part of their psychology and that human agents need a system of self-management in response to the potential of these forms of motivation to conflict with basic commitments. In the absence of such self-management, human agents are much less likely to be effectively guided by their basic commitments.[51]

Once our model of reflective, self-governing agency explicitly includes these further, wide-ranging, first-order motivating attitudes, however, there is pressure for higher-order reflectiveness and conative hierarchy. After all, we can suppose that a self-governing agent will know of these first-order attitudes and of her need for self-management. And we can suppose that she will, other things being equal, endorse forms of functioning that serve this need. So it is plausible to suppose that her basic commitments will themselves include a commitment to associated management of relevant first-order desires and thus include such self-management as part of their content. And that means these commitments will be higher-order. In particular, given the centrality of practical reasoning to self-governed agency, we can expect that these commitments will include policy-like attitudes that concern the justifying significance to be given (or refused) to various first-order desires, and/or what they are for, in her motivationally effective practical reasoning. Such policies will say, roughly: give (refuse) justifying significance to consideration X in motivationally effective practical reasoning, in part by giving (refusing) such significance to relevant first-order

truth about what is good for oneself" (242). Cooper notes that this is compatible with holding, as Aristotle did, that "non-rational desires carry with them value judgments framed in (at least some of ) the very same terms of good and bad, right and wrong, etc., that also reappear in our rational reflections about what to do and why" (247). (In contrast, I would want to allow for some nonrational desires that do not involve such value judgments.) What is central, Cooper indicates, is "the permanence in human beings and the independence from reason . . . of the nonrational desires" (249).

51. For a similar focus on this practical problem—though not in the service of a hierarchical model—see Martha C. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge: Cambridge University Press, 1986): chap. 4. Note that the commitments that need to be supported by self-management will include shared commitments—for example, our shared commitment to a certain project.

desires and/or what they are for (and do this by way of this very policy).[52] Such policies will help shape what has subjective normative authority for the agent.[53]

This means that a basic pressure for conative hierarchy derives from what is for human agents a pervasive practical problem of self-management. In particular, reflective, self-governing agents will have a wide range of first-order motivating attitudes that will need to be managed in the pursuit of basic commitments. This practical problem exerts pressure on those commitments to be higher-order. And once we recognize this point, we can go on to see such higher-order commitments as potential elements in a Frankfurtian project of self-constitution. If, in contrast, we were to try to model reflectiveness, internality, and self-government without appeal to conative hierarchy, we would be in danger of failing to take due account of this pervasive practical problem.

The idea is not that individual agents reflectively decide to introduce conative hierarchy into their psychic economies in response to the need for self-management.[54] Rather, we can agree with Frankfurt that human agents are in fact typically reflective about their motivation in ways that involve conative hierarchy. Our question is: What can we say to ourselves to make further sense to ourselves of this feature of our psychic lives? This question is part of what T. M. Scanlon calls our "enterprise . . . of self-understanding."[55] And the claim is that we can appeal here to the role of higher-order reflection and conative hierarchy as part of a reasonable response to fundamental,

52. See my "Autonomy and Hierarchy." Note that I do not claim that these are the only policies that may be relevant here. For example, as Alfred Mele has noted, the agent may also have a policy in favor of simply trying to remove a certain desire.

53. In including in some such policies a direct concern with *X*, as well as with associated desires and what they are for, I am extending (as anticipated earlier) what it is that is accorded subjective normative authority.

54. Though we, as theorists, can reason in this way, as part of what Paul Grice called "creature construction." See Grice's "Method in Philosophical Psychology (from the Banal to the Bizarre)" (Presidential Address), in *Proceedings and Addresses of the American Philosophical Association* (1974–75): 23–53. I pursue such a methodology in "Valuing and the Will" and in "Autonomy and Hierarchy," In "Autonomy and Hierarchy," I see self-governing policies as a solution to a pair of pervasive human problems: the need for self-management and the need to respond to underdetermination by value judgment.

55. T. M. Scanlon, "Self-Anchored Morality," in J. B. Schneewind, ed., *Reasons, Ethics, and Society: Themes from Kurt Baier with His Responses* (Chicago: Open Court, 1996): 198. As I see it, one use

pervasive, and (following Cooper's Aristotle) permanent human needs for self-management in the effective pursuit of basic commitments.

This is not to argue that self-governance *must* involve conative hierarchy. It is, rather, to argue that there is a pervasive and permanent practical problem that human agents face and with respect to which conative hierarchy is a reasonable and common human response, at least for agents with relevant self-knowledge. The claim is, further, that when the hierarchical response to this pervasive and permanent practical problem takes an appropriate form—one we have tried to characterize—we arrive at basic elements of a central case of self-governance. Because the cited form of hierarchy essentially involves plan-type attitudes—in particular, self-governing policies—we arrive, as promised, at a model of core elements of human autonomy that involves in basic ways structures of planning agency. And because the planning theory is, as I have said, a modest theory of the will, this is a model of central roles of the will in autonomy.[56]

## 8. SOME FINAL QUALIFICATIONS

In discussing Watsonian "perverse" cases, I indicated that self-governance precludes certain kinds of severe incoherence between evaluative judgment and basic commitments. This does not entail that self-governance requires evaluative judgment; nor does it entail that self-governance requires that the agent who does make such evaluative judgments gets them right. Indeed, I think that it is not essential to the basic commitments I have emphasized—those that take the form of self-governing policies and have agential authority—that they derive from intersubjectively accountable value judgments. But it still might be urged that there is a further demand

---

of Gricean creature construction is to help us achieve such self-understanding. Note that in locating this question about conative hierarchy within the enterprise of self-understanding, I do not suppose that the basic concern to which our answer to this question appeals must be a concern with self-understanding. Indeed, the relevant concern to which my answer appeals is a concern with the effective pursuit of basic commitments. For a view that sees this basic concern as, in contrast, a concern with self-understanding, see J. David Velleman, "Introduction," *The Possibility of Practical Reason*, 1–31.

56. These roles are multiple and interconnected: they include the organization of thought and action over time, related forms of agential authority, and roles in shaping what has subjective normative authority. This contrasts with a thin conception of the will as primarily a matter of deciding what to do in present circumstances.

specifically on autonomy, that relevant self-governing policies be to some extent grounded in evaluative judgment—though they may also be underdetermined by, and go beyond, such judgments. And it might also be urged that there is a further demand specifically on autonomy, that the agent at least have the ability to arrive at evaluative judgments that get matters right.[57] These are not, however, issues I will try to adjudicate here. For our present purposes, it suffices to note that whatever we say on these further proposals is compatible with, and could be added to, the proposed model of core elements of autonomy.

Finally, there are traditional and perplexing issues about the compatibility of autonomy and causal determination. The features of agency I have highlighted here as core elements seem to me to be ones that could be present in a deterministic world, which is not to deny that certain forms of causal determination (for example, as the argument frequently goes, certain forms of manipulation) can undermine self-governance. Nevertheless, whether there is a persuasive reason for insisting that autonomy preclude any kind of causal determination of action (because, as the argument might go, causal determination of action is incompatible with self-determination of action) is a matter of great controversy, one that I also will not address here.[58]

57. See, e.g., Susan Wolf, *Freedom within Reason* (Oxford: Oxford University Press, 1990); Nozick, *Philosophical Explanations*, 317–32; and Gideon Yaffe, "Free Will and Agency at Its Best," *Philosophical Perspectives* 14 (2000): 203–29. We need to be careful, though, to remember that our concern here is with autonomy and not directly with moral accountability. [For a related caveat, see Gary Watson, "Two Faces of Responsibility," *Philosophical Topics* 24 (1996): 240–41.]

58. Though see my "Nozick on Free Will."