https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

AUTONOMY AND HIERARCHY*

By Michael E. Bratman

I. The Autonomy-Hierarchy Thesis

In autonomous action the agent herself directs and governs the action. But what is it for the agent herself to direct and to govern? One theme in a series of articles by Harry G. Frankfurt is that we can make progress in answering this question by appeal to *higher-order conative attitudes*.[1] Frankfurt's original version of this idea is that in acting of one's own free will, one is not acting simply because one desires so to act. Rather, it is also true that this desire motivates one's action because one desires that this desire motivate one's action. This latter desire about the motivational role of one's desire is a second-order desire. It is, in particular, what Frankfurt calls a second-order "volition." And, according to Frankfurt's original proposal, acting of one's own free will involves in this way such second-order, and sometimes yet higher order, volitions.[2]

Frankfurt's hierarchical proposal has met with a number of challenges and has been subject to clarification and emendation.[3] I myself have elsewhere tried to map out some details of this debate.[4] My concern here, however, is with the very idea that there is a close connection between autonomous agency and motivational hierarchy.

* Thanks to William Brewer, Alisa Carse, John Fischer, Nadeem Hussain, Margaret Little, Alfred Mele, Elijah Millgram, Henry Richardson, Neil Roughley, Ralph Wedgewood, and audiences at the Georgetown University Philosophy Department and the Social Philosophy and Policy Center, Bowling Green State University. Special thanks to Gideon Yaffe for a series of very helpful discussions. Work on this essay was supported by a fellowship from the John Simon Guggenheim Memorial Foundation.

[1] See Harry G. Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988). See also Gerald Dworkin, "Acting Freely," *Nous* 4 (1970): 367–83; Wright Neely, "Freedom and Desire," *Philosophical Review* 83 (1974): 32–54; and Keith Lehrer, "Freedom, Preference, and Autonomy," *The Journal of Ethics* 1, no. 1 (1997): 3–25.

[2] Frankfurt, "Freedom of the Will and the Concept of a Person," in Frankfurt, *The Importance of What We Care About*, 11–25.

[3] For an important, early response to Frankfurt's original essay, see Gary Watson, "Free Agency," *Journal of Philosophy* 72 (1975): 205–20. Watson offers an alternative approach, one that replaces appeal to motivational hierarchy with an appeal to a distinction between motivational and evaluative orderings. Watson also points to at least two potential concerns for the hierarchical approach: (1) a concern about the grounds for seeing higher-order desires as having a stronger claim to speak for the agent than do lower-order desires, without embarking on an unacceptable regress; and (2) a concern about the idea that, in deliberation, we reflect on our desires rather than directly on our options. I discuss this second concern below, in the main text of this essay.

[4] Michael E. Bratman, "Identification, Decision, and Treating as a Reason," in Bratman, *Faces of Intention* (New York: Cambridge University Press, 1999), 185–206.

 © 2003 Social Philosophy & Policy Foundation. Printed in the USA.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

Of course, much depends on what kind of close connection one has in mind. Some might argue that all cases of human autonomous agency essentially involve motivational hierarchy. But I will focus on a somewhat weaker claim. As I see it, talk of autonomous agency and of autonomous action is talk of a highly abstract property of agents and actions, one that involves agential direction and governance of action. We can ask, what kinds of psychological functioning in human agents are such that they can constitute or realize this abstract property?[5] And we can consider the view that at least one central kind of psychological functioning that can constitute or realize human autonomous agency involves motivational hierarchy. That is, it involves the functioning of higher-order conative attitudes that concern the presence and/or functioning of conative attitudes. Perhaps there are other forms of functioning that could also claim to realize a kind of human autonomy. If there are, then we will want to understand their relation to the hierarchical model. But, at the least, a central kind of functioning that can realize human autonomy involves conative hierarchy. Or so it may be claimed. Let us call this the *autonomy-hierarchy (AH) thesis*. And let us ask why we should accept this thesis.[6]

Gary Watson points to reasons to be skeptical about accepting the AH thesis.[7] Watson notes that 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."[8] It seems to me that Watson is right in arguing that the "initial practical question" that is explicitly and consciously raised in one's practical reasoning is ordinarily about "courses of action" and not about ourselves. But it is one thing to acknowledge this point about the "initial practical question" and another thing to reject the idea that, in at least one central kind of case, autonomy involves motivational hierarchy.

[5] Here I am, broadly speaking, following both Frankfurt and J. David Velleman. See, in particular, J. David Velleman, "What Happens When Someone Acts?" in Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), 123–43. In speaking of functioning that realizes such an abstract property, however, I am making room for the possibility of multiple realizations. I am unsure whether Frankfurt or Velleman would also want to do so. (My appeal in the text to a "central kind of functioning" signals that my concern is with the limited claim that one theoretically important realization involves motivational hierarchy.) Let me also note here that, as I understand the notion of functioning, not all causal impacts will be included in an attitude's functioning.

[6] In Michael E. Bratman, "Reflection, Planning, and Temporally Extended Agency," *Philosophical Review* 109, no. 1 (2000): 35–61, I explore the role, in strong forms of human agency, of higher-order policies concerning the functioning of first-order desires in one's motivationally effective practical reasoning. A number of individuals have asked whether such policies about practical reasoning need to be higher order. (Samuel Scheffler once raised this question in a particularly helpful way in correspondence.) The present essay responds to these concerns.

[7] Watson, "Free Agency," 205–20.

[8] Ibid., 219.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

Indeed, I believe that higher-order conative attitudes play a significant role in central cases of autonomous agency, and so we should accept the AH thesis. In support of this view I offer here two lines of argument. One line of argument (the one that will be my main focus here) derives from the role of valuing in central cases of autonomy, and from pressures on such valuing to involve hierarchy. This argument draws on the idea that an autonomous agent not only governs her actions, but also governs the practical reasoning from which those actions issue. A second line of argument for the AH thesis derives from the idea that an autonomous agent's governance of her own practical reasoning involves her understanding of this reasoning as so governed. In each case, there are reasons to think that a central model of psychological functioning that can at least partly constitute or realize human autonomous agency will make essential appeal to motivational hierarchy. The first step in advancing these arguments is to reflect on the phenomenon of valuing.

II. Valuing and Two Problems for Human Agents

It is sometimes useful in the philosophy of action to see certain features of human agency as (at least, implicit) responses to pervasive and systematic problems that human agents face. I think that this strategy is especially useful when we consider what it is to value something. In particular, I think that we can see valuing as a response to two different, though related, problems that reflective human agents face. I shall describe what these problems are, how valuing constitutes a response to them, and what light this sheds on the higher-order structure of valuing. I shall then explain why it is plausible to see such valuing as central to autonomy.

Many problems that we face as human beings are faced by a wide range of nonhuman agents as well; but some problems that we face are limited to agents who are, like us, reflective in certain ways. Here, I highlight two problems of the latter sort. The first concerns reflective self-management. We are creatures who are affected and moved by complex forms of motivation, and we sometimes find ourselves needing to reflect on, and respond to, these forms of motivation.[9] Suppose that I find myself angry, resentful, and desiring retribution. I am, however, reflective: I ask myself whether, as we say, I "really want" to pursue retribution or, rather, to turn the other cheek. I thereby face a problem of reflective self-management.

The second problem begins to arise once we make judgments of value, judgments that we see as intersubjectively accountable in characteristic

[9] This is a central Frankfurtian theme. The idea of casting this problem together with the problem, noted below in the text, of underdetermination by value judgment parallels aspects of Marth C. Nussbaum's discussion in her *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (Cambridge: Cambridge University Press, 1986), chap. 4.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

ways.[10] On reflection, we can reasonably come to judge that there are many things that have value. We can also reasonably come to believe that a coherent human life frequently involves decisions and/or the assignment of weights, priorities, or other forms of significance that go beyond and are underdetermined by these prior, intersubjectively accountable judgments of value.[11] A dramatic case can be found in a version of Sartre's famous example.[12] A young man sees the value of fighting with the Free French, and he also sees the value of staying with his mother. With respect to these judgments, he may expect an appropriate form of intersubjective convergence. The young man also, however, believes that a coherent, temporally extended life requires some sort of specific, wholehearted commitment to one of these valuable activities over the other, a commitment with respect to which he may well not expect relevant intersubjective convergence. Granted, he may suppose that after he has arrived at a commitment to, say, the Free French, the value of loyalty to his commitment becomes salient; and about *this* value he may expect relevant intersubjective convergence. But this does not undermine the observation that there was underdetermination of the contours of the young man's life by his value judgments prior to arriving at his commitment.

I will call these two problems, respectively, problems of *self-management* and problems of *underdetermination (of the contours of one's life) by value judgment*. These are not only problems that philosophers have in theorizing about human agency. They are pervasive, practical problems faced by ordinary human agents.

This is not to say that these problems are normally ones with which we are explicitly and consciously concerned in our everyday practical thinking. Rather, much of our ordinary, day-to-day practical thinking takes for

[10] In "A Desire of One's Own" (*Journal of Philosophy*, forthcoming), I note several different ways of interpreting this constraint of intersubjectivity. We might, for example, see a judgment of value as made from a Humean "common point of view," or as a judgment that those who are appropriately rational and informed would converge in a relevant way, or as involving the expression of a demand on others to converge in relevant ways. And other interpretations are possible. For our present purposes we do not need to settle on a specific interpretation, though for ease of exposition I will sometimes write in ways that fit most naturally with the second of these interpretations. For a version of this second interpretation see Michael Smith, *The Moral Problem* (Oxford: Basil Blackwell, 1994), 151–77.

[11] A number of philosophers have emphasized ways in which such judgments of value can underdetermine the specific contours of an individual life. For present purposes I will take it for granted, without further argument, that there frequently is some such underdetermination. See, e.g., Isaiah Berlin, *Four Essays on Liberty* (Oxford: Oxford University Press, 1969); Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), esp. 446–50; and Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1986), chap. 14. Consider also T. M. Scanlon's remark that "one cannot respond to every value or pursue every end that is worthwhile, and a central part of life for a rational creature lies in selecting those things that it will pursue." T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 119.

[12] Jean-Paul Sartre, "Existentialism Is a Humanism," in W. Kaufmann, ed., *Existentialism from Dostoevsky to Sartre* (1956; reprint, rev. and expanded, New York: Meridian/Penguin, 1975), 354–56.

granted background structures that help to constitute our solutions to these problems. Watson may well be right in noting that we ordinarily do not reflect explicitly and directly on our motivation. Instead, the direct target of our explicit practical reasoning frequently concerns what to do. Nevertheless, our management of our motivation is one of the problems that needs to be addressed by the structures that help to shape our practical reasoning. More generally, our coordinated responses to problems of self-management and of underdetermination by value judgment are, so to speak, part of the deep structure of our ordinary practical thinking. Or so I propose.

When we see our practical thinking in this way, we can ask the question: What features of such thinking enter into our solutions to the problems of self-management and underdetermination by value judgment? My conjecture is that human agents tend to incorporate into their practical thinking a unified—as it were, simultaneous—solution to this pair of problems. This unified solution is valuing.

III. Valuing and Policies about Practical Reasoning

I propose that human agents tend to incorporate into their practical thinking valuing understood in a certain way. What I have in mind is this: Policies are intentions that are general in relevant ways.[13] We have policies of action. We also have policies, or policy-like attitudes,[14] that concern the significance that is to be given to certain considerations in our motivationally effective practical reasoning concerning our own conduct.[15] I might, for example, have a policy that gives no weight at all to revenge, another policy that gives great weight to family, and yet another policy that gives little or no weight to my own contribution to political goals. On the one hand, such policies partly constitute my stance with respect to relevant motivation, such as a desire for revenge, that might come up for reflective assessment. On the other hand, some policies of this sort constitute my response to the problem of fashioning a life with a coherent shape in the face of underdetermination by value judgment. Such policies, or policy-like attitudes, about practical reasoning are a kind of valuing, one that constitutes a unified response to problems of self-management and of underdetermination by value judgment.

[13] My discussion throughout this essay assumes the approach to intention that I have called "the planning theory" and that I present in Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987; reissued by CSLI Publications, 1999). I discuss policies, esp., at 87–91. I also discuss policies in Michael E. Bratman, "Intention and Personal Policies," *Philosophical Perspectives* 3 (1989): 443–69.

[14] Concerning this qualification see my discussion of what I call "quasi-policies" in Michael E. Bratman, "Reflection, Planning, and Temporally Extended Agency," 57–60. In most of my discussion here I will not keep repeating this qualification (though I will return to it briefly below in note 51).

[15] Cp. Nozick, *Philosophical Explanations*, 446–49.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

I have touched on some of these themes elsewhere. In "Valuing and the Will," I pursue a project of Gricean "creature construction."[16] This project introduces forms of valuing as steps in the "construction" of a series of fictional creatures, in pursuit of a (partial) model of actual human agents. We begin with a creature who has certain broadly conative attitudes—desires in a broad sense—and certain belief-like cognitive attitudes. Early in the project of creature construction, we envision a creature whose desires have been suitably exposed to its relevant beliefs and in that sense are "considered." We then turn to a creature who engages in a primitive form of deliberation, in which its considered desires determine the weight that is given to various factors, where the weight that is given matches the degree to which these considered desires tend to move the creature to action. The considered desires of such a creature can be thought of as a primitive kind of valuing.

A more complex creature, however, might be more reflective about her desires, including her considered desires, and might ask herself how she "really wants" such desires, and what they are for, to enter into her deliberation and motivation. An intelligible output of such reflection would be a higher-order policy, or policy-like attitude, about that creature's treatment of her desires as providing, for her motivationally effective deliberation, *justifying* considerations for action.[17] In "Valuing and the Will," I call such policies *self-governing* policies, and I argue that they constitute an important kind of valuing.[18]

It is helpful, here, to distinguish two different ways in which a first-order desire may enter into practical reasoning.[19] Suppose, for example, that my desire for revenge motivates action by way of associated practical reasoning. In one case the content of my (defeasible) reasoning might be expressed as follows:

[16] The basic idea of creature construction comes from Paul Grice. Grice aimed to "construct (in imagination, of course) according to certain principles of construction, a type of creature, or rather a sequence of types of creature, to serve as a model (or models) for actual creatures." See Paul Grice, "Method in Philosophical Psychology (From the Banal to the Bizarre)," Presidential Address, *Proceedings and Addresses of the American Philosophical Association* 68 (1974–75): 37. My discussion is in Michael E. Bratman, "Valuing and the Will," *Philosophical Perspectives* 14 (2000): 249–65.

[17] For some intermediate steps in this construction see "Valuing and the Will," 252–57.

[18] In the central case that I consider in "Valuing and the Will," the self-governing policy concerns first-order motivation that is already present. I also note, however, that there can be cases in which the policy involves, rather, a commitment to acquiring certain desires; and such a policy might concern one's treatment of certain desires, were one to acquire them. For a related but different conception of a connection between valuing and policies, see David Copp, *Morality, Normativity, and Society* (New York: Oxford University Press, 1995), 177–78.

[19] For a closely related distinction see Philip Pettit and Michael Smith, "Backgrounding Desire," *Philosophical Review* 99 (1990): 565–92. In what follows, my first case corresponds to cases in which, in their terminology, the desire is in the "foreground." My second case is similar to one kind of case in which, in their terminology, the desire is in the "background."

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

Model 1
(a) I desire revenge.
Action *A* would promote revenge.
So I have a justifying reason for *A*-ing.
So I will *A*.

Here (a) is, as is said, the major premise.[20] In a second case, in contrast, my reasoning has as its major premise an appropriate expression of my desire, or of a thought involved in my having that desire. So, for example, we might in the second case see the major premise as:

(b) Revenge is a justifying consideration.

where (b) is an expression of my desire, or of a thought involved in my having that desire.[21] The content of my (defeasible) reasoning would then be along the lines of:

Model 2
(b) Revenge is a justifying consideration.
Action *A* would promote revenge.
So I have a justifying reason for *A*-ing.
So I will *A*.

(Here [b] is understood in the indicated way.)[22]

Now, a self-governing policy that eschews my treating my desire for revenge as reason-providing in my motivationally effective practical reasoning will eschew practical reasoning of both sorts. Indeed, it might do this even in some cases in which my desire really does involve a thought or judgment along the lines of (b). A self-governing policy that supports

[20] We might also see (a) as alluding to further conditions that the desire fulfills, for example, that it is a considered desire.

[21] Appeal to an evaluative expression of the desire is characteristic of Donald Davidson's views about practical reasoning. See, e.g., Donald Davidson, "Intending," reprinted in Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), 85–86. John Cooper emphasizes how, on Aristotle's theory of virtues of character, even appetites and forms of anger and grief involve judgments about the good or what ought to be done, although these judgments are not themselves based on reasoning that aims at determining what is good or what ought to be done. Cooper also emphasizes the permanence of these nonrational desires even in a human being of Aristotelian virtue of character. See John M. Cooper, "Some Remarks on Aristotle's Moral Psychology," reprinted in Cooper, *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory* (Princeton, NJ: Princeton University Press, 1999), 237–52.

[22] We might try to see (b), when it is an expression of (a thought involved in) my desire, as sometimes involving an implicit indexical element:

Revenge is a justifying consideration (from *my* point of view).

We would then need to address the broadly Frankfurtian issue of which point of view is *mine*. This is the issue of agential authority that I turn to briefly below in Section V. A consequence of the approach to agential authority sketched in Section V (see also note 39) is that there are desires that are not appropriately expressed in this way.

my treating my desire as reason-providing in my motivationally effective practical reasoning will support practical reasoning of one or both of these sorts. Note that even in the case in which such a policy concerns only practical reasoning along the lines of Model 2, the policy still concerns the cited functioning of the relevant desire in that reasoning. The policy is a higher-order policy about that functioning of the desire, even though the relevant premise in the policy-supported reasoning—premise (b)—does not itself refer to that desire, but is, rather, an expression of that desire or of an involved thought.[23]

Now, as they emerge from the story of creature construction as so far developed, higher-order, self-governing policies are primarily a response to concerns with reflective management of one's motivational system. In contrast, in "A Desire of One's Own," I highlight not only these issues of self-management, but also the problem that is posed for our agency by our judgments about multiple, conflicting values that, at least so far as we can see, underdetermine what particular, coherent shape our lives are to take.[24] I suggest that our response to this problem will consist, in one important type of case, in policies, or policy-like attitudes, that say what justifying significance to give to various considerations in our motivationally effective deliberations and practical reasoning about our own action.[25] So, to use an example from "A Desire of One's Own," consider a person—let us call her Jones—who sees the value in sexual activity and who also sees the value of a certain kind of life of abstinence. Jones might

[23] Let me note two complexities. The first concerns Model 2. In some cases the desire for *X* will, even prior to an endorsing policy, already involve a thought of *X* as a justifying consideration, or will at least be plausibly expressible along the lines of (b). But there are, I think, also cases which do not fit well into such a picture: for some cases of pre-reflective anger, for example, this will seem to be an overly intellectualistic picture. Nevertheless, if in a case of this latter sort one does arrive at a self-governing policy in support of treating the anger as reason-providing, then this policy may infuse or shape the anger so that it becomes (or involves a thought that is) expressible in this way. So the reasoning supported by the policy can be Model 2 reasoning.

A second complexity concerns motivation in the absence of either kind of practical reasoning. An agent who rejects her desire for revenge has a self-governing policy of not allowing that desire to lead to action by way of Model 1 or Model 2 practical reasoning. I think we can also suppose that the agent's policy rejects an effective motivational role for that desire, even if that role does not involve such practical reasoning—perhaps the desire of a Frankfurtian "unwilling addict" could in some cases motivate action in this latter way. However, it is policies specifically about the roles of desires in motivationally effective practical reasoning that are central to autonomous action; or so I will be claiming below in the text. These policies will be my main concern here.

[24] Bratman, "A Desire of One's Own." For such talk about the "shape" of our lives see Charles Taylor, "Leading a Life," in Ruth Chang, ed., *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press, 1997), 183.

[25] I discuss this idea further in Michael E. Bratman, "Shared Valuing and Frameworks for Practical Reasoning," in R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Oxford University Press, forthcoming). Note that the idea is *not* that such policies directly change what is valuable—though there is room for an indirect impact by way of the value of living in accord with such policies, once they are adopted.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

then arrive, on reflection, at a policy of giving no positive weight to her sexual activity. She sees that there are alternative, nonabstaining ways of living that have value, but in creating for herself a life of abstinence she puts the value of her sexual activity aside, so to speak. Although she does not expect relevant intersubjective convergence on living such a life, she arrives at a policy, or policy-like attitude, concerning her own motivationally effective practical reasoning. And given the role of such a policy (or policy-like attitude) in her practical reasoning and action, it seems reasonable to see it as constituting a kind of valuing.

IV. Two Problems, One Solution?

A salient response to the problem of self-management and to the problem of underdetermination by value judgment involves policies concerning one's own motivationally effective practical reasoning. Such policies say what significance to give to certain considerations in this reasoning. Such policies constitute an important kind of valuing. Valuing in this sense is related to, but is to be distinguished from, judging what is good.[26] Indeed, this distinction lies at the heart of the usefulness of such valuing as a response to the problem of underdetermination by value judgment.

One might, however, question whether this is really a single solution to our pair of problems. After all, although our respective responses to these problems involve policies about practical reasoning, there seems to be a difference in the kind of policy that is cited. The policies that are cited as a response to problems of self-management are primarily higher-order responses to separable forms of motivation: desires for revenge or for sexual activity, for example. The policies about practical reasoning that are a response to concerns about underdetermination by value judgment do not need to be about separable forms of motivation, though they may be. Perhaps in response to his dilemma, the young man in Sartre's example settles on a policy of giving weight to helping the Free French, but not to helping his mother. So described, this policy does not seem to be about the functioning of separable forms of motivation. It seems rather directly to support (defeasible) reasoning along the lines of:

Model 3
(c) Helping the Free French is a justifying consideration.
A would help the Free French.
So I have justifying reason to do *A*.
So I will *A*.

[26] For this distinction see David Lewis, "Dispositional Theories of Value," in Lewis, *Papers in Ethics and Social Philosophy* (New York: Cambridge University Press, 2000), 68–94; Gary Watson, "Free Action and Free Will," *Mind* 96 (1987): 150; and Gilbert Harman, "Desired Desires," in Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2000), 117–36, esp. 129–30.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

At the same time, this policy seems to reject analogous reasoning concerning the young man helping his mother. Here, premise (c)—in contrast with premise (b) in Model 2—need not be an expression of (a thought involved in) a separable desire. So, we might wonder how policies of practical reasoning that are involved in our solution to problems of self-management are related to policies that constitute our solution to problems of underdetermination by value judgment. Do we really have a single solution to our pair of problems?

I think that the basic point to make here is that we need to respond to both problems, and there will be, at the least, a requirement that an agent's responses to these problems mesh with each other. There will be pressure on our young man, for example, not to have policies that give predominant weight to helping the Free French, but that nevertheless encourage the effective influence on his relevant Model 1 or Model 2 practical reasoning of his powerful desire to stay with his mother. In pursuit of a model of autonomy we want a model of a more or less unified agent, one whose agency involves both reflective management of his motivation and a response to underdetermination by value judgment. As reflective human agents, we have both a problem of self-management and a problem of responding to underdetermination by value judgment with a form of, so to speak, limited self-creation.[27] I have been assuming that the latter problem is pervasive. The pervasiveness of the former problem is ensured by the pervasiveness of forms of motivation—including appetites, and forms of anger and grief—that can pose problems of self-management.[28] We seek coordinated solutions to both problems: as we might say, the self that emerges from self-management should be coordinated with the self that emerges from limited self-creation.

We can develop the point further by returning to creature construction. My discussion in "Valuing and the Will" ends with a creature who has self-governing policies concerning which desired ends to treat as justifying considerations in (as I have here described it) her motivationally effective Model 1 or Model 2 deliberation. Such policies play central roles in the organization of the agent's own thought and action over time. They also play central roles in various forms of social organization, coordination, and cooperation. After all, much of our ability to work with and to coordinate with others depends on our grasp of the justifying significance that they give to various considerations in their practical thinking.

This role in social coordination points to the enormous significance of these forms of coordination in the creature's life, a point that Allan Gibbard

[27] For such talk of self-creation see Raz, *The Morality of Freedom*, 385–90; and Joseph Raz, "The Truth in Particularism," in Raz, *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press, 1999), 242–45.

[28] See Cooper, "Some Remarks on Aristotle's Moral Psychology," esp. 247–50, where Cooper highlights the contrast with the Stoics.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

has emphasized with great insight.[29] As Gibbard might say, pressures for social coordination will lead to pressure on our creature to try to articulate, explain, and, to some extent, defend and justify her self-governing policies to others in her social world. This suggests that we can expect to emerge—in a later stage of creature construction—some sort of intersubjectively accountable views about values and/or reasons.[30] But at that point we can also expect that these further views will have a feature highlighted by our second problem: given the need for intersubjective accountability, these views will tend to leave unsettled many questions about the particular contours of an individual agent's life. These views will tend, by themselves, to underdetermine, to underspecify, how one is to live.[31] At least, this is reasonable to expect, given the assumption that such underdetermination is common. So a creature's self-governing policies, formed initially in response to problems of self-management, may be in a position to do "double duty" in this later stage of creature construction. They also may be in a position to help constitute her response to underdetermination by her intersubjectively accountable value judgments.[32] In this way the creature's responses to our pair of problems can be expected to mesh.

We can also consider matters from the other direction, by beginning with policies of practical reasoning that are a direct response to underdetermination by value judgment. Let us here return to Jones. Jones has arrived at a policy of abstinence, a policy that precludes giving positive deliberative weight in her life to her sexual activity. This is her own, distinctive response to underdetermination of the contours of her life by her prior judgments about the good in a human life of sexual activity, on the one hand, and of abstinence as a part of a certain kind of religious observance, on the other hand. Now, this policy seems to be directly about how to weigh certain considerations in her Model 3 practical reasoning, and not about the functioning of separable first-order motivation. But it is likely that in order for this policy to be effective, it will need to involve or be associated with a policy, or policy-like attitude, of putting to

[29] Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990). While I think that Gibbard's focus on issues about social coordination is of great importance, I see my discussion here as neutral concerning the debate between Gibbard's expressivist understanding of value judgment and certain more cognitivist approaches. This is part of an overall strategy—a kind of method of avoidance, to use John Rawls's terminology—of trying to articulate important structures of human agency in ways that are available to a range of different views in metaethics.

[30] I think that we can also expect forms of shared *valuing* (in contrast with shared judgments of value) to emerge. See Bratman, "Shared Valuing and Frameworks for Practical Reasoning."

[31] I think that this is implicit, for example, in Gibbard's effort to distinguish between an "existential commitment" and accepting "a norm as a requirement of rationality." See Gibbard, *Wise Choices, Apt Feelings*, 166–70.

[32] Which is not to say that these self-governing policies may not themselves be responsive to the creature's judgments of value.

one side in her motivationally effective practical reasoning considerations provided by her felt sexual desires. It will need to involve or be associated with a policy of not treating those desires as providing justifying considerations for her Model 1 or Model 2 practical deliberation.

Or consider the young man who settles on a life of fighting with the Free French. This will "mesh" (in the way that I mentioned above) only if he has a way of managing the impact on his deliberation and motivation both of his inclinations not to fight—that is, his affections and concerns for his mother, his fears of battle—and of his affections and concerns for the Free French. A policy of giving weight in his motivationally effective Model 3 practical reasoning to his work with the Free French, but not to his mother's needs for his attention, will likely be effective only if it involves or is associated with such forms of self-management. So there will be pressure on the young man for associated higher-order policies of self-management, policies that concern relevant practical reasoning along the lines of Models 1 and/or 2.

If we begin with problems of self-management, then we arrive first at self-governing policies that are, in part, about the management of the functioning of one's first-order motivations. We arrive later at the idea that such policies can also constitute (part of) a response to underdetermination by intersubjectively accountable value judgments. If, instead, we begin with the latter problem about underdetermination, then we arrive first at policies about the significance of certain considerations in one's motivationally effective practical reasoning, and second at associated higher-order policies about the management of the impact of relevant forms of motivation on one's effective motivation and practical reasoning. I surmise that the differences are not differences in the basic model of human agency, but in our route to that model. At the heart of the model in each case are policies about what is to be given significance in one's motivationally effective practical reasoning. And, given the kind of creatures that human agents are, these policies will normally involve or be associated with policies that concern the management of relevant forms of motivation in practical reasoning and action.[33]

This is not yet to identify these two kinds of policies concerning practical reasoning. The AH thesis does not depend on such identification. The thesis needs only to insist that hierarchical policies are an element in a central case of human autonomy. Nevertheless, I think that there is normally reason for a kind of identification. What we have seen is that in reflective agents like us there is substantial pressure toward a unified

[33] A fuller discussion also would consider both "quasi-policies" (see note 14 above) and "singular commitments." See Michael E. Bratman, "Hierarchy, Circularity, and Double Reduction," in Sarah Buss and Lee Overton, eds., *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge, MA: MIT Press, 2002), 65–85. These complexities can be put to one side here, however, since our primary concern is with a kind of hierarchy involved in all of these phenomena.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

cluster of forms of functioning—a cluster that involves coordinated, cross-temporal, policy-like control of practical reasoning along the lines of Models 1, 2, and 3. This suggests that we see the underlying source of this cluster as a single, complex policy or policy-like attitude. In a central case such a policy will be something like this:

> It will be a policy of giving justifying significance to consideration *X* in motivationally effective Model 3 reasoning, in part by way of appropriate control of associated motivationally effective practical reasoning along the lines of Models 1 and/or 2.

In a central case, this will be the form that a self-governing policy will take. Although the genesis of such a policy might only sometimes include explicitly higher order reflection on first-order motivation, its function and content will be, in part, higher order in the indicated ways.[34] Such (to some extent) higher-order policies are an important form of valuing.

V. Valuing and Autonomy

And they are a form of valuing whose control of action can partly realize or constitute a human agent's direction and governance of action, and thus, the agent's autonomy. Or so I maintain. It is time to say why.

In autonomous action, as I have said, an agent directs and governs her action. Note that there are two different ideas here: agential *direction* and agential *governance*. As I see it, in agential direction there is sufficient unity and organization of the motives of action for their functioning to constitute direction by the agent.[35] Agential governance is a particular form of such agential direction: agential governance is agential direction that appropriately involves the agent's treatment of certain considerations as justifying reasons for action. Autonomous action involves a form of agential direction that also constitutes agential governance. And I want to describe what these phenomena of agential direction and agential governance consist in without appealing to a homunculus account, that is, to a "little person in the head who does the work."

Without appealing to a homunculus account, my strategy is to see agential direction and governance as being realized by appropriate forms of psychological functioning.[36] There is agential direction of action when action is under the control of attitudes whose role in the agent's psychology gives them authority to speak for the agent, to establish the agent's

[34] I consider in the text below, in Section VI, the objection that there may be a gap here between function and content.

[35] See Frankfurt's work on wholeheartedness in, for example, Harry G. Frankfurt, "The Faintest Passion," reprinted in Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), 95–107.

[36] See note 5 above.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

point of view—gives them, in other words, agential authority. This agential direction of action is, furthermore, a form of agential governance of action only when these attitudes control action by way of the agent's treatment of relevant considerations as justifying reasons for action, that is, as having subjective normative authority for her.[37]

When we approach autonomous action in this way, valuing of the sort that we have been discussing seems to be a natural candidate for an attitude whose control of action can, in part, realize the agent's direction and governance of action. On the one hand, self-governing policies play central roles in supporting and constituting important forms of cross-temporal organization and coordination in an agent's life. As long as an agent's self-governing policies are not involved in conflict that undermines these cross-temporal organizing roles, we have reason to see such policies as having agential authority.[38] Hence, we likewise have reason to see their control of action as realizing agential direction of action. On the other hand, such policies function, in particular, by way of helping to shape the agent's operative, background framework of justifying reasons.[39] To borrow terminology from J. David Velleman, the policies' control of action is part of a story not only of motivation, but also of rational guidance.[40] This is why agential direction of action that is realized by the controlling role of such policies constitutes, at least in part, agential *governance* of action. Taken together with the arguments that I have just offered, that such self-governing policies involve (or bring with them) motivational hierarchy, this leads us to the view that motivational hierarchy is at the heart of at least one important realization of human autonomy. This leads us, that is, to the AH thesis.

This argument for the AH thesis has two main steps. The first step is to articulate what we might call design specifications for an autonomous agent.[41] I have, so far, cited two design specifications: sufficient organization of motivation to constitute agential direction, and motivation that involves rational guidance in a way that further qualifies this agential direction as agential governance. The second step in defending the AH

[37] Concerning these two kinds of authority, see Michael E. Bratman, "Two Problems About Human Agency," *Proceedings of the Aristotelian Society* 101 (2001): 309–26.

[38] I expand on these matters, and their relation to ideas about personal identity, in Bratman, "Reflection, Planning, and Temporally Extended Agency." In pages 48–51 of that essay I describe the cited nonconflict condition as a version of what Frankfurt calls "satisfaction." In my discussion of higher-order policies (below in the text) I will take it for granted that some such satisfaction condition is realized. A full account of satisfaction would also need to consider the significance of conflict with singular commitments concerning what to treat as justifying (see note 33 above).

[39] We might say that such self-governing policies help constitute the agent's justificatory point of view. So if such a self-governing policy were to reject a desire for *X*, and that desire were nevertheless to involve the thought that *X* is a justifying consideration from that agent's point of view, that thought would be false.

[40] Velleman, "The Guise of the Good," in *The Possibility of Practical Reason*, 99–122.

[41] Cp. Velleman, "Introduction," in *The Possibility of Practical Reason*, 11.

thesis is to argue that a model in which higher-order self-governing policies function in the indicated ways would satisfy these design specifications.[42]

Must the kind of rational guidance that is needed for autonomy also involve sufficient responsiveness to what is judged to be good and/or is good?[43] My response here is to leave this question open, since either answer is compatible with the present argument for the AH thesis.

A final point on valuing and autonomy is that there might be actions that are not the direct issue of the kinds of policy-directed practical reasoning that I have cited, but that are sufficiently related to such reasoning to be candidates for autonomous action in an extended sense. Once we have in hand our basic model of autonomous agency, we can allow for such extensions in our account of autonomous action.

VI. Two Objections and the Autonomy-Transparency Thesis

I now consider a pair of closely related objections to this way of defending the AH thesis. Each objection acknowledges the role of reasoning-guiding policies in an important realization of human autonomy. Each objection nevertheless goes on to challenge the idea that such policies need be hierarchical.

The first objection concerns self-management. Suppose you find yourself desiring revenge. You stop to reflect, and you arrive at a commitment not to give weight in your practical reasoning to revenge. This commitment, or policy, is certainly a response to an antecedent desire for revenge, but why must its *content* make explicit reference to the functioning of that desire? Why can't its content simply reject forms of Model 3 reasoning that give positive weight to revenge? Granted, for such a commitment to be effective it must somehow involve management of the impact on one's practical reasoning and action of one's desire for revenge. But this does not show that the content of the guiding policy must refer somehow to the functioning of one's desire and cannot simply be the rejection of Model 3 reasoning that gives positive weight to revenge. So it is not clear that what is needed are higher-order policies of the sort highlighted by the AH thesis.

[42] On this account autonomous action is compatible with the persistence of first-order motivation that diverges from what is supported by one's self-governing policies. Within the proposed model, what autonomy requires is that one's self-governing policies actually do guide one's relevant reasoning and action. Further, there can be cases—e.g., our case of principled sexual abstinence—in which one's self-governing policy rejects a desire for *X* even though one acknowledges the value of *X*.

[43] See Susan Wolf, *Freedom Within Reason* (Oxford: Oxford University Press, 1990); but see also Gary Watson, "Two Faces of Responsibility," *Philosophical Topics* 24 (1996): 240. Relatedly, we might also consider a constraint that, at the least, the relevant self-governing policies not favor one's own loss of autonomy or complete domination by others. Here, again, we need not settle the issue in order to argue for the AH thesis.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

The second objection concerns limited self-creation (responding to underdetermination by value judgment). Recall Jones's policy of sexual abstinence, which I discussed in Section IV. I have said that, to be effective, this policy will likely need to involve a policy of eschewing the demands of felt sexual desires on her motivationally effective deliberations. The second objection to the AH thesis grants that for Jones's policy of sexual abstinence to be effective, there will normally need to be management of the impact on her deliberation and action of felt sexual desires. However, this objection claims that it does not follow that the *content* of her policy of abstinence will need to refer explicitly to this management of contrary desires. Her policy can simply eschew appeal in her Model 3 deliberation to her sexual activity as a justifying consideration. So it does not follow that her policy is higher order in its content.

Both objections to the way that the AH thesis has been defended thus far grant that the psychological functioning that issues from the reasoning-guiding policies whose control can partly realize human autonomy will normally include some form of management of the impact of relevant first-order motivation on practical reasoning and action. But both objections insist that it does not follow that the reasoning-guiding policies must themselves be higher order in their content.

Now, we might simply respond that there are cases and cases. As long as there are common cases in which autonomy is realized by the functioning of reasoning-guiding policies that are higher order, the AH thesis stands. But I think that we can make a stronger claim than that here.

Our concern is, after all, with *autonomous* agency. For such agency, some sort of modest condition of *transparency* seems apt in characterizing the relation between, on the one hand, known significant functioning that is supported by a reasoning-guiding policy and, on the other hand, the content of that policy. We can put the idea this way:

> If one knows that the effective functioning of the practical reasoning that is supported by one's reasoning-guiding policy at the same time significantly involves management of the roles of relevant motivation, and if one's ensuing action is self-governed, then the content of one's reasoning-guiding policy will refer to, and support, this management of motivation.

The idea here is that, in the absence of such transparency, the functioning of the reasoning-guiding policy would not be sufficient to ensure an agent's governance of his or her ensuing action. Let us call this the *autonomy-transparency (AT) thesis*. The AT thesis helps to block the current pair of objections. These objections depend on driving a wedge between the psychological functioning that issues from the reasoning-guiding policy and the content of that policy. The AT thesis blocks this wedge for cases of autonomy in which the agent has the requisite self-knowledge.

And since the requisite self-knowledge need only be a fairly general knowledge of the need for management of motivation—a kind of self-knowledge that is compatible with only partial knowledge of the specific complexities of one's actual motivation—it seems plausible to suppose that an autonomous agent will be knowledgeable in this way.

VII. Transparency and Self-Governed Practical Reasoning

I believe that, if accepted, the autonomy-transparency (AT) thesis effectively blocks the cited two objections to the autonomy-hierarchy (AH) thesis, but why should we accept the AT thesis? My answer appeals to pressures on an autonomous agent to govern not only her action but also the practical reasoning from which her action issues.[44]

Let us begin by noting that the fact that there is practical reasoning leading to action, and that this reasoning has normative or evaluative content, does not yet ensure that the agent governs the reasoning. There can be cases of motivationally effective practical reasoning about which we will want to say—borrowing a phrase from Gibbard—that the agent is not governing the reasoning but is instead in the "grip" of concerns that drive the reasoning.[45] This might happen, for example, in a case involving a strong desire for revenge and associated thoughts about what degree or type of revenge is deserved. One's motivationally effective practical reasoning might be in the grip of this desire and these normative thoughts.

So the model of (a central case of) autonomous agency that we have been developing—a model that appeals to the role of self-governing policies in guiding reasoning and action—needs to include psychological functioning that ensures that the agent is not in the grip of relevant concerns but is, rather, governing her own, relevant practical reasoning. Suppose, then, that relevant practical reasoning, and its control of action, involves in an important way psychological functioning of type *F*. And suppose that the agent does govern that reasoning and knows that it involves this *F*-type functioning. We can expect that the agent's governance of her reasoning will extend to that known functioning. And the natural way, within the model, to ensure agential governance of that *F*-type functioning is to build support for such functioning into the content of a self-governing policy that guides her reasoning. And this is a condition that will be violated when the agent's reasoning is, instead, in the grip of a certain consideration.

[44] As I understand her views, this is a theme in Christine M. Korsgaard's *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), chap. 3. It appears here in my discussion as, in effect, a third design specification on autonomous agents.

[45] Gibbard, *Wise Choices, Apt Feelings*, 60. I discuss the significance of such cases also in Bratman, "Hierarchy, Circularity, and Double Reduction" and in Bratman, "Two Problems About Human Agency."

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

Return now to the AT thesis. This thesis, which I described at the end of the preceding section, concerns cases of self-governed actions that are the issue of practical reasoning that is guided by a relevant self-governing policy, *P*. According to the AT thesis, if the agent in such a case knows that the effective functioning of that practical reasoning significantly involves management of her relevant motivation, then the content of *P* will refer to and support this management of motivation. We have now observed that if, in such a case, the action is self-governed, then so is the practical reasoning from which it issues. And we have also observed that if an agent is to govern her practical reasoning in such a case, then she must govern known, important *F*-type functioning involved in that reasoning and its control of action. We have, further, provided a model of psychological functioning that would realize such agential governance of that reasoning. And within this model, the agential governance of this *F*-type functioning will involve guidance by a policy whose content supports this *F*-type functioning. The AT thesis is, then, a special case of this general feature of this model of self-governed practical reasoning—a special case in which *F* is the management of the cited functioning of relevant motivation. This means that at least a central case of autonomous agency will involve the kind of transparency that is needed to complete our first argument for the AH thesis.

Related concerns about transparency also suggest a second line of support for the AH thesis; or so I now proceed to argue.

VIII. The Second Line of Argument: Policies about Self-Governed Practical Reasoning

According to the model that we have been developing, an autonomous agent's reasoning-guiding policies guide practical reasoning that is, in part because of this guidance, governed by the agent. That this reasoning is agentially governed is, I take it, something that the autonomous agent will normally know and endorse. But then a natural extension of our reflections on transparency suggests that the agent's guiding policy will be a policy that favors practical reasoning that is governed by herself. Within the model, however, in order to be governed by the agent the reasoning needs to be guided by a relevant self-governing policy. So it will be plausible to expect that the reasoning-guiding policy is, in part, about its own role in guiding the reasoning. This is to build into the content of the policy that guides the practical reasoning the condition that this same practical reasoning be appropriately guided by that very policy. And this is to draw on work by Gilbert Harman and others on forms of *reflexivity in intentions*.[46]

[46] Gilbert Harman, "Practical Reasoning," in Harman, *Change in View* (Cambridge, MA: MIT Press, 1986), chap. 8; and Harman, "Desired Desires." See also Alan Donagan, *Choice:*

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

Harman, in particular, has argued that a "positive" intention in favor of an action will be "an intention that something will happen in a way that is controlled or guided by" that very intention. In this sense, a positive intention "is reflexive or self-referential—it refers to itself."[47] Harman cautions that this does not require that the agent have "an explicit mental representation of her intention."[48] What is required, however, is that what the intention favors is that there be a certain process that is suitably dependent on, and responsive to, that intention itself.

Consider, then, a case in which one knowingly governs practical reasoning in which one gives positive weight to revenge. The idea now is that one's policy of giving weight in one's reasoning to revenge will be a policy that one's reasoning give such weight to revenge, in part, *because of* this very policy. It will be a policy of giving such weight as a matter of this very policy.

This suggests that, at least in certain central cases of autonomous action, the self-governing policies that guide the underlying practical reasoning will be reflexive;[49] that is, they will be in part about their own guidance of the practical reasoning.[50] Such a reflexive self-governing policy will be a higher-order conative attitude. It will be a policy about the functioning in reasoning of a certain policy, namely, itself. So we arrive again, as promised, at a form of motivational hierarchy.

Note, however, that this form of motivational hierarchy is different from that at stake in our first line of argument. According to our first line of argument, valuing involves policies that are, in part, about the role of desires, and/or of what they are for, in providing justifying premises in motivationally effective Model 1 or Model 2 practical reasoning. According to the second line of argument that I have just sketched, the practical reasoning of an autonomous agent, at least in a central case, involves policies that are, in part, supportive of their very own functioning in guiding practical reasoning. The functioning of these self-governing pol-

The Essential Element in Human Action (London: Routledge & Kegan Paul, 1987), 88; John Searle, *Intentionality* (New York: Cambridge University Press, 1983); and J. David Velleman, *Practical Reflection* (Princeton, NJ: Princeton University Press, 1989). For an important critique of these ideas, see Alfred Mele, *Springs of Action: Understanding Intentional Behavior* (New York: Oxford University Press, 1992), chap. 11; Harman's response is in Harman, "Desired Desires."

[47] Harman, "Desired Desires," 121. Let me note that I am not here endorsing Harman's general view that all positive intentions are reflexive. I am only using his idea of reflexive intentions to make progress with the special case of self-governed practical reasoning.

[48] Ibid., 124. Harman notes here a parallel with John Perry's observation that (as Harman writes) "a child can have the thought that 'it is raining' without having any concepts of places or times and without any inner mental representations of particular places and times, even though the content of the child's thought concerns rain at a particular place and a particular time."

[49] Related ideas about reflexivity can be found in Keith Lehrer, *Self-Trust: A Study of Reason, Knowledge, and Autonomy* (New York: Oxford University Press, 1997), 100–102.

[50] I offer a related argument for seeing such policies as reflexive in Bratman, "Two Problems About Human Agency," 323.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

icies, which is reflexively supported in this way, is their very own guidance of practical reasoning along the lines of any of Models 1 through 3.

IX. Conclusion

Recall Watson's observation that in normal cases of practical reasoning "the initial practical question" is "about courses of action and not about ourselves." My defense of the autonomy-hierarchy (AH) thesis is consistent with this view. My claim is not about the initial practical question, but about the background structures that are brought to bear in trying to answer this question. I have argued that, in central cases of autonomous action, these background structures involve higher-order self-governing policies. While we frequently take for granted such structures in our practical reasoning, we have seen reason to think that they are present in at least certain central cases of autonomy.

Granted, I have left open the possibility that other kinds of background structures might also satisfy our design specifications for autonomy. Faced with such an alternative proposal, we would want to see whether our design specifications are indeed satisfied, and, if so, whether there are, at bottom, significant similarities with our model of self-governing policies.[51] But such prospects can be left open here.

Now, recall that Frankfurt's original appeal to motivational hierarchy—to what he called higher-order "volitions"—was an appeal to higher-order conative support for the functioning of a first-order desire as an effective motive of action.[52] I have been led here to higher-order policies not only in support of forms of functioning (along the lines of Models 1 or 2) of first-order motivation in one's practical reasoning and action, but also in reflexive support of their own framework-providing role. Both of these forms of policy-supported functioning in practical reasoning go beyond the bare motivational role of first-order desires, which is the concern of (at least, the original version of) Frankfurtian higher-order volitions. Nevertheless, the approach to autonomy that has emerged here shares with Frankfurt's approach the basic idea that some hierarchical structures provide an important element of at least one central case of autonomous human agency.

The AH thesis is a thesis about important kinds of contemporaneous psychological functioning that can partly realize human autonomous agency. Certain issues about the history of elements in this functioning remain open. In particular, it may be that, in the end, a full story about human autonomy will also need to appeal to some sort of historical condition that blocks certain extreme cases of manipulation, brainwash-

[51] And, if so, whether the idea of a quasi-policy can usefully capture these similarities.

[52] Frankfurt, "Freedom of the Will and the Concept of a Person," 16. But see the modification of this idea in Harry G. Frankfurt, "Identification and Wholeheartedness," in Frankfurt, *The Importance of What We Care About*, 159–76.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press

ing, and the like.[53] This is not an issue to be settled here.[54] But before we can settle this issue, we need the best account available of the structural conditions involved on the occasion of autonomous action. My concern here has been to argue that our account of such structural conditions should endorse a version of the autonomy-hierarchy (AH) thesis.

Philosophy, Stanford University

[53] See John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998); and Keith Lehrer, "Reason and Autonomy" (in this volume). See also Bratman, "Fischer and Ravizza on Moral Responsibility and History," *Philosophy and Phenomenological Research* 61, no. 2 (2000): 453–58. Note, though, that the present issue is autonomy, not the related but different idea of moral responsibility. (See Gary Watson's distinction between "attributability" and "accountability" in Watson, "Two Faces of Responsibility.")

[54] Of course, if the specification of the content of the relevant attitudes is ineluctably historical (for reasons developed by, among others, Tyler Burge and Hilary Putnam), then we would need to appeal to such content-fixing historical considerations. But that is a different matter.

https://doi.org/10.1017/S0265052503202065 Published online by Cambridge University Press