https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

THE NORMATIVITY OF SELF-GROUNDED REASON*

By David Copp

I. Introduction

In this essay I propose, and then attempt to ground, a standard of practical rationality. According to this standard, to a first approximation, rationality consists in the efficient pursuit of what one values.[1] This standard differs from the familiar principle of instrumental reason, which requires us to take the most efficient means to our ends, for it gives special emphasis to those of our ends that qualify as our values. It also differs from the principle of self-interest, which requires us to pursue our own good, both because we might value the good of others as much as our own good, and because, if we are unwise, we might value things that are bad for ourselves. I speak of the conception of rationality I develop as "self-grounded" because it requires the pursuit of a person's *own* values, and also because, as I shall argue, a person's values are grounded in her *identity,* on one useful conception of the identity of persons.

The idea of *grounding* a standard of rationality will require some discussion. The term comes from Immanuel Kant, as does my strategy, broadly understood, for I aim to ground the standard of self-grounded reason in the idea of autonomous agency, which of course is a strategy inspired by Kant's work in the *Groundwork of the Metaphysics of Morals.*[2] The account

* I am grateful to many people for helpful comments and discussion over the many years in which I have been developing the ideas in this essay. With apologies to those whose help escapes my memory, I would like to thank Nomy Arpaly, Sam Black, Michael Bratman, Justin D'Arms, Dan Farrell, Pat Greenspan, Don Hubin, Dan Jacobson, Marina Oshana, Michael Ridge, Michael Robins, David Sobel, Pekka Väyrynen, and David Velleman. I presented early versions of some of the ideas in this essay to audiences in the departments of philosophy at the University of Alberta, the University of Maryland at College Park, l'Université de Montréal, the University of Southern California, and the University of Florida, to the 1999 Conference on Moral Theory and Its Applications, Le Lavandou, France, and to the 2001 Conference on Reason and Deliberation, Bowling Green State University. I am grateful for the helpful comments of those who participated in the discussions on all of these occasions and especially to the other contributors to this volume, and its editors. I owe special thanks to Ellen Paul for encouraging me to integrate my thinking on identity with my thinking on rationality and for her useful comments.

[1] This formulation ignores a qualification I have discussed elsewhere. See David Copp, *Morality, Normativity, and Society* (New York: Oxford University Press, 1995), chap. 9; and David Copp, "Rationality, Autonomy, and Basic Needs," in Neil Roughley, ed., *Being Humans* (Berlin: de Gruyter, 2001), 334–55.

[2] Immanuel Kant, *Groundwork of the Metaphysics of Morals,* trans. James W. Ellington (Indianapolis: Hackett, 1981). See p. 50 (Ak 448). In references to the *Groundwork,* I refer first to the pagination of the Ellington translation and then to the pagination of the Prussian Academy Edition (Berlin: de Gruyter, 1902–).

© 2005 Social Philosophy & Policy Foundation. Printed in the USA.

I shall give of autonomous agency is very different from Kant's, however, and my conception of rationality is modest by comparison with a Kantian conception.[3] Most importantly, it does not guarantee a strong Kantian link between rationality and morality, for a person can exhibit a rational efficiency in the pursuit of what she values even if her values are morally abhorrent. Moreover, the grounding that I shall propose is compatible with naturalism; it does not rest on claims that purport to be synthetic a priori, nor does it depend on a nonnaturalistic metaphysics.

The basic idea is this: To be autonomous is to be *self*-governing. To be rational is at least in part to be self-*governing;* it is to do well, by a standard that we need to specify, in *governing oneself.* I argue that a person's values are aspects of her *identity* in a way that most of her ends are not, and that it therefore is plausible to view action governed by one's values as *self*-governed. This is also plausible on independent grounds. Given this, I say, rational agents comply with a standard—the "values standard"—that requires them to serve their values, and to seek what they need in order to continue to be able to serve their values.[4] I argue, then, that there is *reason* for an agent to serve her values and to seek what she needs in order to continue to be able to serve her values. An agent's values are a source of reasons—reasons of a kind that any rational person would take into account in deliberation, if she were aware of them, just in virtue of being rational. I call such reasons "self-grounded reasons," or "reasons of autonomy." I sometimes call the conception of self-grounded rationality the "autonomy conception."

The autonomy conception combines an account of the *content* of the standard of rationality with an account of its *grounding*. I begin, in Section II, by outlining some of the advantages of the conception. In Sections III and IV, I explain the conception of autonomous agency, and I briefly discuss Michael Bratman's model of such agency. In Section V, I develop an account of the identity of persons, and in Section VI, I propose a conception of an agent's values and link it to my account of identity and through it to the conception of autonomous agency. I then turn to the content of the standard of rationality. In Section VII, I outline the values standard, and, in Section VIII, I address an important objection, the objection from ungrounded ends. In Sections IX and X, I explain the idea of

[3] Christine Korsgaard has recently suggested that "the normativity" of the principle of instrumental rationality "must be traced to the agent's self-government, specifically to his capacity to be motivated to shape his character in accordance with an ideal of virtue." See Christine Korsgaard, "The Normativity of Instrumental Reason," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon Press, 1997), 220 n. 13. My account does not invoke a moralized conception of autonomy.

[4] In speaking of "serving" a value, I mean to cover both cases of taking means that are instrumental to fulfilling or realizing something one values and cases of doing what is constitutive of fulfilling or realizing a value. For this usage, see Gavin Lawrence, "The Rationality of Morality," in Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn, eds., *Virtues and Reasons* (Oxford: Clarendon Press, 2002), 90.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

grounding a standard and argue that the values standard can be grounded in the conception of autonomous agency. Finally, in Section XI, I explain the way in which, on the autonomy conception, rationality has priority for practical deliberation.

II. Why Take This Road?

I believe that the autonomy conception has both intuitive and theoretical advantages over the familiar alternatives. One advantage is that it is compatible with a pluralistic view of reasons and normativity, a view that could be called "reasons pluralism." I do not aim to defend reasons pluralism in this essay, and an advocate of the autonomy conception could consistently deny it. Nevertheless, I think reasons pluralism is plausible. To focus the discussion, I offer an example.[5]

Imagine a group of mountain climbers who, after weeks of struggle, are within a day's climb of the summit of Everest. They will have to turn back tomorrow. Just after they begin their final push for the summit, they come across a small party of climbers who are huddled together, clearly in terrible need of help. These people, "the victims," are in their predicament as a result of an extraordinary combination of circumstances, including unusually extreme weather. They need to be helped down to a lower elevation, and their needs are immediate and life-threatening. The first group might realize that, morally, they ought to help. Yet they are indifferent to moral considerations; they do not care whether they are doing what they morally ought to do. They hurry on to the summit. In doing so, they are being extremely selfish and callous, but, intuitively, this is compatible with their being entirely *rational.*

This example brings out several intuitive points about our ideas of rationality and reasons for action, points that are accommodated by the autonomy conception.

First, the example suggests that a person can be rational in knowingly doing something morally wrong. Kant would disagree, I take it, but I shall not argue against the Kantian view in this essay.[6] My focus will be on developing the autonomy conception rather than on arguing against alternatives or exploring the relation between rationality and morality.

One might think that if the climbers believe they ought morally to help the victims, then they have moral values, and thus, on the autonomy conception, they have a self-grounded reason to help. I will argue, however, that an agent's values (in the relevant sense) are not beliefs. Despite their beliefs, the climbers may lack moral values, and thus, on the autonomy conception, they may be rational to ignore any temptation to help

[5] The example is based on an event described in Jon Krakauer, *Into Thin Air: A Personal Account of the Mount Everest Disaster* (New York: Anchor, 1998).

[6] Kant, *Groundwork,* 50–51 (Ak 447–49).

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

the victims. Rationality does not ensure moral virtue. A rational person may have moral values, but she need not.

Second, I think the example supports the intuitive plausibility of reasons pluralism—the view that there are different kinds of reasons, including moral reasons, self-grounded reasons, reasons of etiquette, and so on. And the example suggests, third, that if rationality is a matter of responsiveness to reasons, it is a responsiveness to reasons of a certain kind. Even if there are reasons of etiquette, for example, a rational person might decide not to give them any weight, just as, in the Everest example, the climbers decide to set aside moral considerations. On the autonomy conception, the reasons that a rational agent takes into account in deciding what to do, if she is aware of them, simply in virtue of being rational, are self-grounded reasons—these are (roughly) facts about the impact the agent's alternatives would have on what she values.[7]

Fourth, in the example it seems intuitively plausible both that the climbers rationally ought to continue the climb and that they morally ought to help the victims. Intuitively, just as there are different kinds of reasons, there are different kinds of "ought." This raises the question of what the climbers ought to do *simpliciter.* I have argued elsewhere that there is not a highest-order normative standard or kind of reason relative to which there is in general something that ought to be done *simpliciter.*[8] If I am correct, there is no answer to the question of what the climbers ought to do *simpliciter,* but I will not be arguing for this position here. The autonomy conception is compatible with the thesis that the rational action is required *simpliciter,* but does not entail it.[9]

One might object that when we are making a decision, we want to know what to do *period,* not merely what to do *rationally.* My account speaks to this concern, but not by assigning a special *metaphysical* status to self-grounded rationality. Rather, as I shall explain, it assigns a *deliberative* priority to self-grounded rationality. Self-grounded reasons play a central role in rational deliberation that is not played by other kinds of reasons, such as moral reasons—unless the agent has the relevant values. One might insist in light of this that, on my account, self-grounded rea-

[7] Philippa Foot rejects reasons pluralism in "Morality as a System of Hypothetical Imperatives," in Philippa Foot, *Virtues and Vices* (Berkeley: University of California Press, 1978), 161 and 168 n. 8. Yet she appears to agree with me that moral considerations are not a source of reasons of a kind that any rational person would take into account in deliberation, if she were aware of them, just in virtue of being rational.

[8] David Copp, "The Ring of Gyges: Overridingness and the Unity of Reason," *Social Philosophy and Policy* 14, no. 1 (1997): 86–106.

[9] There is risk of terminological confusion. Some writers think that the English expression "ought" refers, when unqualified, to what I would say "ought rationally" to be done. For example, where Gavin Lawrence speaks of "what the agent ought unsubscripted (that is *qua* rational) to do," I speak of what an agent "ought rationally" to do. See Lawrence, "The Rationality of Morality," 120. I think there is no answer to the question of what the climbers ought to do *simpliciter,* but I do think that, in the example as I understand it, they ought *rationally* to continue the climb.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

sons are the only *genuine* or genuinely *normative* reasons. But there are moral considerations that count for and against our decisions, and it would obscure this fact if we denied that moral reasons are "genuine" or "normative." But this is a side issue. It concerns how best to formulate reasons pluralism, not how best to formulate the autonomy conception.

The autonomy conception has certain theoretical advantages in addition to its intuitive advantages. First, it seeks to *ground* the values standard. Donald Hubin has proposed that rationality consists in conformity to something like the values standard, but he sees no need to ground the standard.[10] Admittedly, it will not be clear why it is an advantage to provide a grounding until I have explained what this involves, but the objective is to explain the normativity of the values standard.[11] It is to explain why so-called self-grounded reasons deserve the label "reasons." I agree, then, with Christine Korsgaard that a principle of practical reason "needs a normative foundation."[12] Second, my account offers a kind of realism about rational requirements that is compatible with metaphysical naturalism. On my account, thoughts about rational action and choice are beliefs with naturalistic truth conditions.

In short, I think that the autonomy conception is at the core of an intuitively plausible view about reasons and rationality, which can be integrated into a plausible overall view of deliberation and choice. In the Everest example, if we think the climbers were acting rationally in pressing on to the summit, it is because we understand them as efficiently pursuing their underlying values. Of course, our concern with our *own* rationality is not due, at least not primarily, to a concern to *understand* what we are doing. It is due to a concern, inter alia, to do well in governing our lives.[13] The autonomy conception can explain this. It conceives of rationality as consisting in a kind of success in serving one's values, and as I conceive of valuing, it is partly constitutive of having values (in the relevant sense) that a person have certain policies for her own behavior. On this picture, as I will explain, rational behavior instantiates a kind of self-government.

III. Autonomous Agency

The literature on autonomous agency is primarily concerned to explain how we can be self-governing despite the variety of causal influences

[10] See Donald C. Hubin, "The Groundless Normativity of Instrumental Rationality," *Journal of Philosophy* 98 (2001): 445–68.

[11] There are different conceptions of normativity. I discuss some of the varieties in David Copp, "Moral Naturalism and Three Grades of Normativity," in Peter Schaber, ed., *Normativity and Naturalism* (Frankfurt: Ontos-Verlag, 2004), 7–45.

[12] Korsgaard, "The Normativity of Instrumental Reason," 218, 249.

[13] David Velleman proposes that "self-understanding" is the "constitutive aim" of action—in something like the way in which "truth" is the "constitutive aim" of belief. David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), 16–24, see 22. My account does not require postulating a "constitutive aim" of action. A thorough discussion of Velleman's proposal is beyond the scope of this essay.

on our actions and despite the possibility of causal determinism. The explanandum in this literature is free intentional action. My project is a different one. I shall set aside the worry about determinism. Moreover, the explanandum in my project is, crudely, intentional action that is *governed* by the agent in a way that she "identifies" with. It is action that is autonomous in a different, "thicker" sense. Cases of *akrasia* or weakness of will count as autonomous action in the thinner sense, if any actions do, but in general they do not count as autonomous in the thicker sense that concerns me, for an *akratic* agent normally does not identify with her action in the relevant way. I shall focus on the thicker notion.

Autonomy is a matter of being "self-governing." Consider the idea of a self-governing country or state. States are affected in many ways by the actions of other states and by the environment in which they operate, but excessive outside interferences of certain kinds are incompatible with self-government. In a self-governing state, law and public policy originate in decisions of the government, and the government is able to implement its decisions without being subject to certain kinds of interference. By analogy, an autonomous person is someone capable of deciding for herself what to do, and capable of executing these decisions without interference of certain kinds. To be autonomous, one must meet conditions of two kinds. There are "internal" conditions, including the requirements of being able to make decisions, to form intentions, and to act on one's intentions, and there are "external" conditions, including the requirement of being free of certain kinds of interference, including coercion and manipulation.

The external requirements of autonomy do not play a role in my account of rationality, for one can deal rationally with external interferences, such as coercion and manipulation. The autonomy conception explains rationality in terms of the internal requirements of self-government; thus, for my purposes here, we can set aside the idea that there are also external requirements.[14]

The internal requirements of autonomy are psychological and physical properties and capabilities. First, one must have a will. That is, one must have a structure of beliefs, values, and desires, and the ability to decide how to act and to form intentions to act on the basis of these beliefs, values, and desires. Second, one must have the ability to make one's will effective in leading one to act. That is, one must have the ability to act on one's decisions, to act intentionally. One must have the power to perform bodily movements that, if one is successful, will constitute doing what

[14] Our values are shaped by the values of our family and culture, but such influences do not undermine our autonomy. Were it not for such influences, we might not be capable of the kind of planning required for autonomy. See Will Kymlicka, *Multicultural Citizenship* (Oxford: Oxford University Press, 1995), chap. 5, esp. 82–84. See also Marina Oshana, "Personal Autonomy and Society," *Journal of Social Philosophy* 29 (1998).

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

one has decided to do. The notion of intentional action needs to be explicated, but I believe it is clear enough for my purposes. Autonomous behavior is intentional, in that the agent could in principle give her reason for acting as she does, where her reason is a function of her intention in acting. Of course, intentional action can be irrational, so it is not necessary that the agent's reason for acting be a good reason.

There are important distinctions to be drawn among conative states. Most important for my purposes is the distinction between ordinary desires and intentions. I might desire to eat some ice cream, even though I intend not to do so, and I might intend to do something, such as to adhere to rigid guidelines about grading, even though, in the ordinary sense of the term, I have no desire to do so. Michael Bratman has proposed that intentions are "planful states" in a way that desires are not.[15] The key point, however, is that there is a distinction.[16]

There is also an important distinction between "intentions in action" and "prior intentions."[17] A prior intention can be a specific intention—an intention to do some relatively specific kind of thing at a specific future time. But some intentions govern a kind of action in kinds of situations that I might find myself in on many occasions. An example is my intention to wear a seatbelt when driving. Bratman suggests that we think of general intentions of this kind as "policies." The idea can be usefully extended to intentions that can be achieved only if a plan is implemented over time, where this involves acting on various more specific intentions that are components of the plan. Policies are intentions that are functionally general in that they guide the formation of various specific intentions over time.[18]

To understand the thicker notion of self-governing agency, we need to focus on the idea of *governing* something, understood as a matter of regulating the thing, or exercising systematic control. Autonomous agency is agency that is controlled or regulated by the agent.[19] The ability to have policies or plans is necessary to self-governing agency, so understood; and to the extent that we are self-governing, we shape our lives in accord

[15] Michael Bratman, *Intentions, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

[16] Korsgaard has suggested that "willing an end just is *committing* yourself to realizing the end" (Korsgaard, "The Normativity of Instrumental Reason," 245). She understands such commitment normatively, as analogous to "making a promise" (ibid., 245 n. 60). In my view, to will an end is to form a kind of intention, which is not essentially a normative matter. Korsgaard's view is disputed in R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," *Philosophers' Imprint* 1 (2001): 3–10. http://www.philosophersimprint.org/001003/.

[17] John Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983), 84–85.

[18] Gideon Yaffe pressed me to explain the sense in which policies are general.

[19] For a valuable discussion of the agential control of action, see John Martin Fischer and Mark Ravizza, S.J., *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press, 1998).

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

with our policies or plans. At a minimum, we are not simply driven by our strongest desires, but we select which desires to satisfy and which to treat as ends. There is, for example, the experience, when standing on a tall observation platform, of feeling drawn to jump. To the extent that we are autonomous, we can ignore such desires; we can refrain from treating their objects as goals to achieve. This ability to decide which of our desires to act on is part of what is involved in the ability to plan.[20] Beyond this, in planning, we decide among alternative future courses of behavior, settle on priorities and strategies for achieving our priorities, and of course, if we are self-governing, we act on the basis of such strategies and plans. A person may have various plans or policies, and qualify as acting autonomously in the thin sense, even if she fails in various ways to control or regulate her action systematically on the basis of her plans or policies. She might follow her plans only in a haphazard manner; her decisions might not tend to serve her plans well, or they might not reflect the priorities she has settled on in her planning, and so on. Hence, an agent can be autonomous in the thinner sense without being self-governing in the thicker sense that interests me.

In this thicker sense, I shall argue, a self-*governing* person regulates or controls her actions on the basis of intentions and plans that serve her values, or are at least constrained by her values, such that serving those intentions and plans does not conflict with serving her values. I shall argue that our values are a central subset of our policies, but this is a different point. Here the point is that the values of an autonomous agent constrain the rest of her policies and goals—the ones that do not qualify as values. Hence, the policies and goals that shape her decisions have been shaped by her values. Of course, autonomous agents can act on urges and desires, but they indulge such things within boundaries set by their values. In what follows, when I speak of autonomous agency, I shall intend self-governing agency in this thick sense.

IV. Michael Bratman on Autonomous Agency

In introducing the thick notion of autonomous agency, I described it as agency that is controlled and regulated by the agent herself in a way she "identifies" with. I have not yet explained in what sense an autonomous agent identifies with the way she controls actions that are governed by her values. Before proceeding to explain this, it will be useful to consider a different model of self-governing agency, the model that has been pro-

[20] Michael Bratman stresses the ability we have to decide whether to take the object of a desire to be an end. See his "Identification, Decision, and Treating as a Reason," *Philosophical Topics* 24 (1996): 1–18. Bratman cites Rachel Cohon, "Internalism about Reasons for Action," *Pacific Philosophical Quarterly* 74 (1993): 265–88.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

posed by Michael Bratman.[21] His model is similar to mine, and my proposal builds on his work in the theory of action. The main difference between our views is that Bratman invokes a metaphysical conception of the identity of persons in order to explain the sense in which self-governing behavior issues from the self. I invoke, instead, a nonmetaphysical idea, the idea that an agent can "identify" with a way of controlling action.

Bratman aims to provide a model of the "core elements of autonomy" in at least a significant family of cases of autonomous agency, which he calls cases of "hierarchical self-governance."[22] His model agrees with mine in seeing autonomous action as regulated by a subset of the agent's policies. But whereas I think the relevant policies are our *values*, in Bratman's model they are "self-governing policies."[23] These are second-order policies concerned with the functioning of desires and other conative states in practical reasoning. An example would be my policy of giving weight in deliberation to my desire for safety. Such policies function to guide deliberation. Moreover, because they typically are stable across time, and because guidance by them involves reference to plans and desires the agent has at other times, they play a role in organizing the agent's life across time. They play this role by means of "continuities and connections" of the kind that, as Bratman reminds us, are central to Lockean accounts of personal identity.[24]

I need to explain this. At any time that a person is conscious, she is having various experiences, thoughts, feelings, emotions, and so on. As time passes, the person has new experiences, remembers past experiences, anticipates the future, regrets the past, forms plans, adopts goals, learns new things, and so on. These events in the psychological life of a person are related to one another in a variety of ways that go beyond their mere ordering in time. We are not dealing with a kaleidoscopic flux of unrelated events. For instance, there are memories of previous experiences and anticipations of future experiences; there are plans that are formulated at one time and carried out in specific intentions at a later time. A person might have forgotten many childhood experiences by the time she is in middle age, but as a teenager she likely still remembered many of them, and in middle age she might remember much of what occurred to her as a teenager. In this way, events in childhood might be linked to the memories of middle age by an overlapping chain of memories. Given all of this, the Lockean idea is that a person *is* essentially a

[21] Bratman presents his account in a series of papers. A useful overview is in Michael Bratman, "Planning Agency, Autonomous Agency," in James Stacey Taylor, ed., *Personal Autonomy* (Cambridge: Cambridge University Press, 2004), 33–57.

[22] Ibid., 35–36.

[23] Ibid., 43.

[24] Ibid., 41–42. Bratman refers to Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 206–8, for a contemporary discussion of a Lockean view of personal identity.

stream of psychological events and states that is unified by the fact that the events and states in the stream are linked together in a chain by the kinds of psychological continuities and connections to which I have been referring.

We can now return to Bratman's view. His idea is that self-governing policies function to guide deliberation in a way that involves reference to plans and desires the agent has at other times, including especially future times. These policies therefore organize the agent's life across time by means of continuities and connections of the kind that, in Lockean accounts of personal identity, organize what would otherwise be merely a kaleidoscopic sequence of psychological events into a unified life of a single person. Because of this, Bratman argues, self-governing policies are fitted to constitute "the agent's practical standpoint."[25] They are fitted to do this provided the agent is "satisfied" with them. Bratman explains the latter idea negatively: to be satisfied with a self-governing policy P is for P *not* to be in conflict with one's other self-governing policies (and quasi-policies) in a way that tends to undermine the role of P in supporting Lockean continuities and connections.[26] Finally, Bratman points out, the policies that figure in his model will be self-referential because they will speak to their own functioning.[27] In sum, Bratman holds that, "in a basic case," self-governing agency consists in "the known guidance of practical thought and action by [reflexive] self-governing policies with which the agent is satisfied."[28]

This is an important and elegant model of self-governing agency, and it deserves a sustained discussion beyond what I can give it here. But I have two worries.

First, I believe it is a mistake to think that autonomous action is typically guided by self-governing policies. I think that what is crucial to self-government is that one's life be governed by one's *values*. In order to make it fully clear why I say this, I need to develop my accounts of values and identity, which I will do in the next two sections of the essay. But the basic point is simple. I value my safety, for instance, and values of this kind are not self-governing policies because they are "first-order." In valuing my safety, I am concerned directly with my safety, not with my conative states or with my practical deliberation. It seems to me, however, that actions that are controlled or governed by first-order values of this kind have at least as good a claim to qualify as autonomous as do actions that are controlled or governed by Bratman's self-governing policies. Most

[25] Bratman, "Planning Agency, Autonomous Agency," 42.

[26] For Bratman, a "quasi-policy" is a "higher-order, policy-like" concern that is not, strictly speaking, a self-governing policy. Michael Bratman, "Reflection, Planning, and Temporally Extended Agency," *Philosophical Review* 109 (2000): 57–60 and 49–50. See also Bratman, "Planning Agency, Autonomous Agency," 44, 38–39.

[27] Bratman, "Planning Agency, Autonomous Agency," 44.

[28] Ibid.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

of the values of a typical person are not self-governing policies, since they are first-order. It seems to me that self-governance by such values is typical of autonomous agency.

In response, Bratman could point out that seeming first-order values often consist in clusters of policies that include self-governing policies.[29] For example, valuing my safety might involve a policy of not permitting my love of adventure to outweigh safety in my deliberation. Still, presumably I would have this policy because I value safety. In general, it seems plausible that self-governing policies are adopted to serve our values. So I see no reason to privilege self-governance by self-governing policies by comparison with first-order policies, such as the simple policy of seeing to my safety. Hence, I think Bratman's model describes a special case of autonomy rather than the typical case.

Second, I think we need a richer account of the agent's endorsement of, or identification with, the elements of her practical standpoint. "Satisfaction" with reflexive self-governing policies is inadequate. To see this, imagine an obsessive person who is obsessed with giving no weight in practical deliberation to her obsessions (including this one). This obsession appears to qualify as a self-referential self-governing policy, and the person may count as "satisfied" with it since its Lockean role may be unimpeded by conflict with any other self-governing policies (or quasi-policies).[30] If so, then it may qualify as part of the person's practical standpoint. This seems to be a mistake, because the person neither endorses this obsession nor endorses its having a role in her deliberation. If it has such a role, she might be ashamed that it does. After all, she is obsessed with giving it *no weight.* So I think we should exclude it from her "practical standpoint."

It seems to me, then, that there are two main problems with Bratman's account. In my view, autonomous action is, roughly, action guided by a person's values, where, as I will explain, a person's values figure in her "identity" in a sense that ensures that she endorses their governing her action. A person's values can be first-order, so my view avoids the first problem with Bratman's account. Moreover, as I will explain, the fact that a person's values figure in her identity ensures that they are relevantly endorsed, so my view also avoids the second problem.

[29] Bratman urged this point in a helpful correspondence about my criticisms of his view. See also Michael Bratman, "Autonomy and Hierarchy," *Social Philosophy and Policy* 20, no. 2 (2003): 160–63.

[30] Her other obsessions may conflict with it, but they typically would not be self-governing policies (or quasi-policies). An obsession with her hair, for example, would be first-order, so it would not be a self-governing policy or quasi-policy, and it might not affect her deliberation, for her obsession not to give weight in deliberation to her obsessions stands in its way. If so, it would not impede the Lockean role of the latter. On Bratman's account, she counts as satisfied with the latter provided that no self-governing policy or quasi-policy undermines its Lockean role. See note 26 above.

V. Autonomy and the Identity of Persons

The life of an autonomous agent is governed in some important sense by the agent *herself.* This may suggest that an adequate account of autonomy needs to invoke a metaphysical conception of what agents *are*—a metaphysical conception of the *identity* of persons. As we saw, Bratman invokes a Lockean conception of the person. I think, however, that an account of autonomy should be neutral among various metaphysical accounts of the person. What we need, to use Bratman's term, is a viable conception of the agent's "practical standpoint," which can be understood, roughly, as the set of psychological states of the person that, when they control her behavior, qualify that behavior as *self-governed action.* The idea of a practical standpoint presupposes that we can identify a person in order to draw a distinction within the person's psychology between different ways that action is controlled, so it assumes that the metaphysics of persons is not at issue. The issue we face is psychological, not metaphysical.

In this section, I shall propose an idea of the psychological identity of persons, and use it to develop a conception of the agent's practical standpoint. The basic idea is that certain of a person's beliefs about her life constitute a whole that we can call her "identity" because of the way that these beliefs ground emotions of esteem, such as pride and shame. Such emotions are related to a person's conception of herself because she "identifies" in a relevant way with their objects. Philippa Foot has written: "The characteristic object of pride is something seen . . . as in some way a man's own."[31] I think that a corresponding point could be made about shame and other emotions of esteem. I propose, then, to take a person's identity—her "self-esteem identity"—to be constituted by (roughly) the set of propositions about her life that she believes and that ground emotions of esteem.[32] A person's practical standpoint can be seen as that part of her identity that is concerned, inter alia, with plans for her life—plans which are such that her beliefs about her success or failure in accomplishing them ground emotions of esteem.

The use of the term "identity" to express a psychological notion of the sort I have in mind is fairly recent.[33] Korsgaard has proposed that a

[31] Philippa Foot, "Moral Beliefs," in Foot, *Virtues and Vices,* 113–14. Foot's dictum would need to be reformulated to deal with cases in which pride takes a propositional object, for a proposition is not viewed as "a man's own." A person may be proud of a painting, or proud *that* he made the painting. Following Foot, we could perhaps say, in cases of the latter kind, that the object of a man's pride is a proposition about an actual or possible state of affairs involving the man—where, she would add, the obtaining of that state of affairs is viewed as some sort of achievement or advantage.

[32] I develop this suggestion in detail in David Copp, "Social Unity and the Identity of Persons," *Journal of Political Philosophy* 10 (2002): 365–91.

[33] The 1933 edition of *The Oxford English Dictionary* (Oxford: Clarendon Press, 1933) does not give an entry for "identity" with anything like the relevant meaning. In the 1976 *Supplement* to the O.E.D., however, the term "identity crisis" is given, with the first cited usage being 1954. Perhaps, then, we should speak of the "baby boomer conception of identity."

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

person's identity is a system of characteristics that the person has and values having.[34] But I want to allow for cases in which a person disvalues an aspect of her identity. During apartheid in South Africa, blacks were abused and humiliated on account of being black, and it is likely that many of them disvalued being black. Yet being black is a property that I would want to treat as part of their identity. K. Anthony Appiah has suggested that the identity of a person is a set of "properties important for social life," which might "matter to their bearers in very different ways."[35] But a person might be tormented by an apparently trivial event that she takes to be quite central to her life even though, and perhaps because, it was not important socially. I would want to treat this as an aspect of her identity.

We should evaluate accounts of identity on the basis of their explanatory usefulness. Different accounts might have different explanatory merits. I initially developed the idea of self-esteem identity in order to cast light on some issues in political philosophy, especially the power and importance of nationalism and patriotism. In this essay, I will use the idea to illuminate the difference between self-governed action and action that is intentional but that conflicts with our values or is governed only by relatively shallow and perhaps transient goals or desires.

I have suggested that the life of an autonomous agent is governed by her own policies and plans, and, more specifically, by her *values,* which are a subset of her policies and plans. The connection with the idea of identity is, I shall argue, that a person's values qualify as aspects of her identity. This remains to be explained, but if we assume it to be correct, then we can see a connection between governing one's life on the basis of one's values and governing one's life *oneself.* I have introduced the idea that an agent who regulates her life on the basis of her values qualifies as self-*governing.* I am now suggesting that she also qualifies as *self*-governing, for her values figure in her identity and constitute her practical standpoint. They ground emotions of esteem that reveal the shape of her self-conception. Moreover, I want to argue, *any* self-governing agent governs her life on the basis of her values. For otherwise, given that her values are aspects of her self-esteem identity (as I will explain), she would be ashamed or disappointed in herself on account of her actions, which would indicate a failure in self-government as she herself sees things. I want to say, then, that an agent is self-governing just in case she governs her life on the basis of her values. Self-governed behavior is not merely intentional and uncoerced. It is regulated by our values, and given that our values are deep psychological features of ourselves that affect our fundamental atti-

[34] See Christine Korsgaard, *The Sources of Normativity,* ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996), 101. Korsgaard speaks of our "practical identity."

[35] K. Anthony Appiah, "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction," in Charles Taylor et al., *Multiculturalism: Examining the Politics of Recognition,* ed. Amy Gutmann (Princeton, NJ: Princeton University Press, 1994), 150–51.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

tudes toward ourselves, and thereby figure in our identity, self-governed behavior also in this way *expresses* the agent's *identity.* The argument turns on the idea of self-esteem identity, which I shall now proceed to explain.

I need to begin with a brief discussion of the concept of self-esteem.[36] Self-esteem is a matter of the degree to which one feels satisfied with oneself on balance, and this is a matter of the degree to which one has a sense of *worth.* A sense of worth should not be identified with a set of beliefs about one's value. A person might believe herself to be valuable, or even to be superior to others, but have low self-esteem, feeling unworthy and insecure. Another person might believe herself to be mediocre, but, despite this, have a solid sense of self-esteem. Self-esteem involves an emotional assessment of oneself, or an emotional stance toward oneself.

A range of emotions is involved. Call them "emotions of esteem."[37] On the positive side, a person can feel good about herself, satisfied or comfortable with herself, or have a sense of worth or security or confidence in herself. A person can take pride in various things to which she takes herself to be related in a relevant way. A person can feel enhanced by something. On the negative side, a person might feel worthless or dissatisfied or uncomfortable with herself, or have a sense of insecurity, or lack confidence. She can feel shame, humiliation, or embarrassment. She can feel diminished by something. All of these emotions can bear on a person's self-esteem.

A person's emotions of esteem are grounded in her beliefs, often in a cluster of beliefs, in one of two ways. First, an emotion of esteem might take as its object a proposition that the person believes. For example, a person might be ashamed that she stole a radio. Second, the person might have a belief she would cite, or the propositional object of which she would cite, to explain her feeling. For example, a person might explain being ashamed that she stole the radio by remarking that she knew at the time she was doing something wrong.

Emotions of esteem can be fleeting. It might be a temporary and short-lived fact about me that I feel ashamed of myself for leaving a miserly tip for a waiter. But even if the shame I feel about the tip is short-lived, it might be a relatively enduring fact about me that I feel ashamed when I recall gaffes I have committed, such as leaving a small tip. We are inter-

[36] In this and the following six paragraphs, I follow the argument in David Copp, "Social Unity and the Identity of Persons."

[37] Gabriele Taylor speaks of "emotions of self-assessment" in her essay "Shame, Integrity, and Self-Respect," in Robin S. Dillon, ed., *Dignity, Character, and Self-Respect* (New York: Routledge, 1995), 168. Taylor holds that self-esteem is primarily to be understood in terms of pride and humiliation (p. 173). See also John Deigh, "Shame and Self-Esteem," in Dillon, ed., *Dignity, Character, and Self-Respect.* Deigh relates self-esteem to a range of emotions of the kind I mention (pp. 135–39). He links the idea of having disgraced oneself and the idea of shame with the idea of who one is, and he links the ideas of disgrace and shame to self-esteem.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

ested in stable and enduring facts of this kind about a person's emotions of esteem and their grounds.

Consider, then, the set of propositions about a person that are believed by the person and that ground emotions of esteem in her in one of the two ways given above, in a stable or relatively enduring way. Let us provisionally define a person's self-esteem identity as consisting in this set of propositions. Suppose, for example, that you are proud that you have Greek ancestry, ashamed that you cannot speak Greek, embarrassed that you have a tendency to leave miserly tips, and mortified that you stole a radio. In this case, your belief in certain propositions explains these emotions (or the propositions are the objects of the emotions). These propositions are included in your identity.

We need to amend this account, for there are cases in which people would feel emotions of esteem in various hypothetical circumstances that seem diagnostic of their identity. Imagine a man who is not proud of his Greek ancestry, but who would feel diminished if he somehow came to believe that he did *not* have Greek ancestry. If this is true of him, his having Greek ancestry ought to be counted as part of his identity. There are also cases in which a person's emotions of esteem would be affected if, counterfactually, she believed certain things about other people, or about a group or an entity, to which she takes herself to be relevantly related.[38] For example, the man might feel pride if someone else whom he believes also to be Greek won the Nobel Peace Prize. In this case too, having Greek ancestry ought to be counted as part of the man's identity.[39]

Here, then, is the proposal.[40] The self-esteem identity of a person at a particular stage in her life is the set of propositions about herself, each of which she believes, where her belief grounds an emotion of esteem. In some cases, her belief "actively" grounds such an emotion. In other cases, her belief grounds such an emotion "potentially" in one of two ways. Either it *would* ground an emotion of esteem if she had certain relevant beliefs about other people or about a group or entity to which she takes herself to be relevantly related, or, if she came to believe its negation, this new belief would ground an emotion of esteem. More formally:

> The proposition that she has property F or that she is R-related to an entity E is an element of S's identity during a stage *s* of her life just in case S believes the proposition during *s* and *either*
>
> (a) this belief grounds an emotion of esteem in a stable and relatively enduring way during *s*, or

[38] For this purpose, we consider the closest possible world in which the person has the beliefs in question. I am assuming the account of counterfactuals in David Lewis, *Counterfactuals* (Oxford: Blackwell, 1973).

[39] Nomy Arpaly helped me with this kind of example.

[40] This paragraph is taken with some modifications from my "Social Unity and the Identity of Persons."

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

(b) it would do so, if S had certain beliefs about E or about other people whom she believes to be F, or to be R-related to E, or
(c) if S were to come to believe during *s* that she is not F, or that she is not R-related to E, then, other things being equal, this belief would ground an emotion of esteem in a stable and relatively enduring way during *s*.

For present purposes, the important point is the connection between self-esteem identity and self-government. Suppose that you have planned your life around various projects, such as raising a family, excelling in your career, being fair in your dealings with others, and so on. The degree to which you are content with yourself will be grounded, among other things, in the degree to which you believe you are finding success in these projects. Since self-contentment is an emotion of esteem, this means that the fact that these are your projects and that you are succeeding in them is an aspect of your identity. Moral commitment and moral character are also entangled with your identity. If a person subscribes to a moral principle, attitudes of shame, guilt, or contentment would be grounded in her record of compliance with the principle. Indeed, the fact that she subscribes to the principle likely would be an aspect of her identity, for she likely would feel ashamed if she came to believe that she does not actually subscribe to the principle. And the things a virtuous person believes about her moral character also normally ground emotions of esteem such as pride or shame, either actually or potentially. A virtuous person would see herself as honest, as not manipulative, and so on, and she would feel ashamed or diminished if she came to see herself as dishonest or as manipulative. Hence, her identity would normally include propositions such as that she is honest and nonmanipulative.

Similar considerations show that a person's *values* are aspects of her self-esteem identity. If a person values honesty, she would tend to feel ashamed of herself if she were to realize she had acted dishonestly, for instance, and she would also feel ashamed of herself if she came to believe that she is not an honest person. Hence, the fact that she values honesty will be an aspect of her self-esteem identity. Nor is this merely likely to be the case. It seems to me that a person who values honesty must have a tendency to feel ashamed if she thinks she has acted dishonestly or if she comes to believe she actually is not honest. This is why the presence or absence of such emotions is evidence of a person's values.[41] In general, then, on my account of valuing, and on my account of identity, a person's values are aspects of her identity.

There is a technical difficulty that needs to be addressed. A person's identity is a set of propositions she believes. The difficulty is that, on my

[41] Aristotle suggests that our character is revealed in our feelings. See Aristotle, *Nicomachean Ethics,* book 2, chapter 3, 1104b3ff. I owe this reference to Elaine Sternberg.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

view, as I will explain more fully in the next section, a person's values are policies or general intentions rather than beliefs. Moreover, such a policy can be merely implicit in the sense that one might not be able to formulate it and one might not follow it self-consciously.[42] Hence, if I want to maintain that a person's values are "aspects" of her identity, it appears I need to say what proposition a person believes when she values V such that (a) it qualifies as an element of her identity on my account of self-esteem identity, and (b) it concerns V in a relevant way or is appropriately related to V. Suppose, then, that you value honesty. In this case, I think you would believe that you have a policy of being honest. This belief might be merely implicit, in that you might not have it consciously in mind; and it might be rather inarticulate, in that, for example, you might not think of yourself as having a "policy" rather than simply a desire to be honest.[43] Nevertheless I think some such belief would be present. Jay Wallace has argued that our intentions are ordinarily "fairly accessible to consciousness," given the role they play in shaping our deliberation. He argues, in fact, that someone who does not believe that she intends to do x "cannot really be described as having the intention to do x."[44] The point, of course, is that policies are general intentions. My proposal, then, is that if a person values V, the proposition that she has a policy of pursuing V is an element of her identity. Indeed, a cluster of beliefs regarding V will normally be included in her identity, including especially beliefs about her success or failure in serving V. For example, if she believes she missed an opportunity to serve V, she will tend to be disappointed in herself. A person's values are revealed by the emotions of esteem that she feels or tends to feel, and these values are embedded in her psychology in virtue of their connection with such emotions. Given all of this, I think it is appropriate to speak of a person's values as being "aspects" of her identity.

A person *endorses* the role of her values in governing her actions in that she is content with herself when she deliberates and acts in accord with them, is ashamed to fail to do so, and so on. To see the importance of this, notice that a person with an obsession might be ashamed of her obsession and ashamed to indulge it. In such a case, the fact that she has the obsession would be an aspect of her identity, but we would not want to count indulging it as an instance of self-government. This is for two reasons. First, the obsession is not an aspect of the person's identity in the

[42] We presumably have a policy of speaking grammatically, but we might not be able to formulate the rules of grammar and might not follow the policy self-consciously. The policy is "implicit" in this sense. Frank Jackson describes grammar as an "implicitly understood theory." See Frank Jackson, "Cognitivism, A Priori Deduction, and Moore," *Ethics* 113 (2003): 569.

[43] Jay Wallace speaks of beliefs that are not "explicitly and articulately present to the consciousness of the agent" but that "are implicit in the agent's understanding of their situation." Wallace, "Normativity, Commitment, and Instrumental Reason," 21 n. 52.

[44] Ibid., 22.

right way. Control of one's deliberation and actions by one's values is control by an aspect of one's identity that is endorsed in one's identity, where such control is also endorsed in one's identity. Obsessions are not normally endorsed in this way, nor is their role in one's life. Second, an obsession is not normally a policy, and the compulsive nature of an obsession subverts the control of one's actions by policies. In discussing Bratman's view (in Section IV), I gave an example of an obsessive policy, but the policy in that example was in part a policy of giving itself no weight in deliberation, so the agent did not endorse its playing a role in her deliberation.

Let me then return to the argument that began this section. If an agent regulates her life on the basis of her values, she qualifies as *self*-governing, for her values figure in her identity. Her actions express her self-conception. And she qualifies as self-*governing*, for she endorses control of her actions by her values and endorses her values. Moreover, if an agent is self-governing, she governs her life on the basis of her values, for otherwise she would be ashamed or disappointed in herself on account of her actions, which would indicate a failure in self-government as she herself sees things. Hence, as I claimed, an agent is (thickly) self-governing just in case she governs her life on the basis of her values. Her values constitute her practical standpoint.

VI. Values as Policies for Action

I have argued that (thickly) autonomous agency is agency that is regulated by the agent's values. And I have suggested that our values are a kind of policy, where a policy is a kind of general intention. I now need to defend this suggestion. I suggested that our values figure in our *identity* in a way that ordinary policies and plans need not. I want to argue that, because of this, and because we endorse control of our actions by our values, to the extent that we are *self*-governed, our actions are shaped and constrained by our values. To make the argument go through, however, I need to show that values can be understood as things that *can* regulate behavior, and to achieve this, I want to argue that values are a kind of policy for action. I also need to argue that our values are basic or fundamental policies that, to the extent we are self-governed, constrain the more occasional and instrumental plans and ends that shape action. The fact that our values are *intrinsic* policies helps to support this view, as I shall explain, as does the fact that our values figure in our *identity* in a way that ordinary policies and plans do not. The latter fact means that there are emotions of esteem to back up the controlling and regulating function of our values. It means that the engine of self-esteem helps to regulate the actions of a self-governing agent.

Let me begin to explain why I hold that our values—or our values "for action"—are a kind of policy. We need to distinguish, of course, between

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

the things that a person values and her state of valuing those things. Consider, for example, what would be involved in valuing honesty, friendship, and so on. I think it is plausible that a person who values honesty must have a policy of being honest. She might sometimes fail to be honest, of course, but honesty must be her policy. It is not enough, for example, that she be honest merely as a result of finding honesty in her best interest from time to time. T. M. Scanlon suggests that to value friendship is among other things to have a policy of being a "good friend," of being loyal to friends, of being concerned with their interests, of spending time with them, and so on. It is also to have a policy of seeking to have friends, to keep the friends one has, and to want those we care about to have friends of their own.[45] Thus, I think that to value honesty and friendship would be, inter alia, to have a set of rough-and-ready policies regarding truth-telling, spending time with friends, and so on. In general, I think, to value something "for action" is, inter alia, to have a kind of policy with respect to it.

More is involved in valuing than having policies. The policies in question must be relatively stable. Our values do not switch on and off like lightbulbs. They are policies of and for temporally extended beings. Normally, moreover, a person who values something is satisfied with the relevant policy and does not regret it or desire to lose it.[46] Of course, we can imagine circumstances in which a person who values honesty would regret her honesty. Perhaps honesty has cost her a friend, for example. A person can have a kind of alienation from her values, perhaps viewing herself as naive for valuing honesty the way she does.[47] Nevertheless, such regret or alienation cannot go too far. For if a person values honesty, she would tend to feel shame or guilt or disappointment in herself, or regret, if she failed to be honest on a given occasion. Some of our policies are not like this. For example, even if I have a policy of reading the newspaper every morning, I would not feel disappointed in myself if, on some occasion, I listened to the news on the radio instead. This shows, I think, that I do not *value* reading the newspaper as such. Perhaps I value getting the news, but if so, then I would feel disappointed in myself if I failed to do anything at all to get the news. Our values for action, therefore, are part of our self-esteem identity, as I have already explained.

I now want to add that I have in mind things that we value *intrinsically.* Our values are "intrinsic policies." This idea is easiest to characterize negatively, as follows. If a policy is intrinsic, we do not have it merely because we think that carrying it out will be or may be instrumental to carrying out other policies that we have or achieving other things that we

[45] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1988), 88.

[46] Here I am influenced by Bratman's views about self-governing policies, in his "Planning Agency, Autonomous Agency," 41–45.

[47] Kadri Vihvelin suggested that a person can be alienated from her own values.

want.[48] Instrumental plans and ends rest on other ends, but intrinsic policies do not. For example, my policy of wearing a seatbelt is not intrinsic, since I have that policy only because complying with it contributes to my safety. But my policy of seeing to my safety is intrinsic. I do not have it merely for instrumental reasons. Thus, given that our values are intrinsic policies, they are more basic in the government of our lives than our more occasional and instrumental plans and ends.

I propose, then, that for a person to value something "for action" is at least in part for her to have an intrinsic policy of choosing or acting in relevant ways, a policy that is relatively stable, that on the whole she is content to have, a policy compliance with which affects her emotions of self-esteem and which is therefore partly constitutive of her identity, and a policy whose role in governing her actions she endorses.[49] This account could be taken as a stipulative, but I think it meshes with an ordinary understanding of what is involved in valuing.

One might object that there are cases in which valuing does *not* seem to involve having policies for action, at least not centrally. These are cases in which we value things we believe we cannot affect in any significant way. For example, I might value the accomplishments of Aristotle.[50] I realize I can do nothing to affect the accomplishments of Aristotle, so my valuing them would seem not to involve my having any policies.

In response, I propose that, even in cases of this kind, a person must have relevant *dispositions.* Suppose, for example, that a person values the pristine environment of Ellesmere Island. If she came to know that the island is threatened by pollution, she surely would have a tendency to support calls for conservation and the like. Having such dispositions is the kind of thing involved, inter alia, in having a policy of the relevant kind. Moreover, the problem cases seem to be ones in which a person values an instance of a kind that she values, where valuing the *kind* involves having a relevant policy. For example, a person who values Aristotle's accomplishments may value them as an example of philosophical accomplishment, and if so, she would have corresponding policies, such as a policy of encouraging, supporting, and applauding such accomplishments. But the important point is that there is such a thing as having a policy, and valuing something "for action" does involve having relevant policies.

My proposal, after all, is to explain rationality and self-grounded reasons in terms of the values standard, and the values standard is to be

[48] Compare Scanlon, *What We Owe to Each Other,* 79.

[49] In *Morality, Normativity, and Society,* I explained valuing in a slightly different way, in terms of subscribing to a standard, but there is a close connection between subscribing to a standard and having a policy. See 177–78, 84, 87–88. I proposed a different kind of account in "Reason and Needs," in R. G. Frey and Christopher Morris, eds., *Value, Welfare, and Morality* (Cambridge: Cambridge University Press, 1993), 112–37. It was criticized in Gilbert Harman, "Desired Desires," ibid., 138–57.

[50] Various people pressed me to discuss examples of this kind.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

understood as concerned with our values for action, values the having of which involves, inter alia, having relevant policies. Our "values for action" are a subclass of our values. (If this is doubted, as I said, we could take my usage to be stipulative. The values standard would then be concerned with policies of the relevant kind.)

Various other proposals have been made about what is involved in valuing something. Scanlon holds that to value something is to take oneself to have *reasons* for holding positive attitudes toward it and for acting in certain ways in regard to it.[51] There are several problems with this proposal. First, it makes valuing something more of an intellectual matter than seems plausible, since it seems to imply that one must have the concept of a reason in order to value anything. Second, it seems that a person's values might run contrary to her beliefs about reasons. A person might take herself to have (aesthetic) reasons to hold positive attitudes toward listening to classical music, but she might not value listening to it. She might not enjoy it, and she might have a policy of avoiding it.[52] If she enjoyed jazz and listened to jazz on a regular basis, then it would be more revealing, I think, to describe her as valuing jazz rather than classical music, regardless of whether she would agree that she has (aesthetic) reasons to listen to jazz. Finally, and most important, Scanlon's proposal is not open to me. He aims to explain valuing in terms of taking oneself to have a reason, but one thing I am trying to explain is what a person would be thinking, in thinking she had a reason. I need an account of valuing that does not use the notion of a reason because my goal is to explain self-grounded reasons in terms of the attitude of valuing.

One might suggest that to value something is to *believe* it is valuable. It seems to me, however, that this proposal also makes valuing more of an intellectual matter than is plausible. A person might have no *beliefs* about the value of listening to music yet still *value* listening to jazz. The proposal also runs into difficulty with pluralism about the bases of value. A person might believe that listening to classical music is aesthetically valuable yet think that listening to jazz is valuable as a source of pleasure. Her values might track the latter beliefs rather than the former. Finally, a person's values might run contrary to her beliefs about what is valuable. She might believe that listening to classical music is valuable, yet she might strongly dislike it and have a policy of seeking out jazz. We *could* say that she "values" listening to classical music, simply to report her belief, but in most contexts, I think it would be misleading to say this.

It is plausible, nevertheless, that valuing something (in my sense) typically goes hand in hand with believing the thing is valuable or good, in at least some respect. Such belief is not necessary, however. In many cases, valuing something (in my sense) *precedes* believing that the thing is valu-

[51] Scanlon, *What We Owe to Each Other,* 95; see also 87–100.
[52] This example was suggested to me by Richard Schubert.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

able. For example, in most cases, I believe, people come to value having a family without first having beliefs about the value of a family. There is perhaps a biological basis to our valuing having a family, which would help to explain this. For my purposes the important point is that, as the example suggests, a *belief* in the value of something is not invariably prior in the order of explanation to *valuing* the thing (in my sense). Moreover, the example shows that it need not be *irrational* for a person to value something in a case in which the fact that she values it (in my sense) is not explained by a belief that the thing is valuable. And it shows as well that it need not be irrational to value something (in my sense) *without* believing it is valuable. We are not irrational to value having a family in the years before maturity brings an understanding of why having a family is valuable.

One might object, however, that it *would* be irrational to value something (in my sense) while believing that the thing has *no* value *at all* or that it is not good in *any* respect. It might seem that there would be a kind of incoherence in this combination of valuing and belief. A person is content with herself when she acts in accord with her values, but it might seem that it would be incoherent to be content with oneself for acting in accord with a policy if one believes that it, or its object, has no value at all. In response, I concede that emotions of esteem depend on one's beliefs in subtle ways.[53] Yet, despite this, a person can value something that she believes not to be good in any respect, and can do so, I think, without irrationality. Someone might value listening to jazz even if she believes there is nothing good at all about listening to jazz. When asked why she spends so much time listening to jazz, she might reply that she loves it even though she can't explain why. I am not convinced that there would be an incoherence in her state of mind or an irrationality in her behavior.[54]

My proposal, then, is that to value something is, inter alia, to have a policy. Bratman has suggested that to have a policy is to have a general intention governing kinds of action in kinds of situations.[55] If this is correct, it is no surprise that deliberation that begins with our values can lead to action. If I have a policy of seeing to my safety, and if this involves having a general intention, then we can see why I form an intention to wear a seatbelt on getting into a car since I believe that wearing a seatbelt serves my safety. If to value something is to have a general intention, then in reasoning about how to achieve what we value, we will form more

[53] This idea has been addressed by many writers on the emotions. See Foot's brief groundbreaking discussion in her essay "Moral Beliefs," 113–14.

[54] This point needs more discussion. Steven Wall helped me to see the need to mention it.

[55] Having such a policy could be understood as a matter, inter alia, of intending to comply with certain relevant standards or norms. This suggestion brings my discussion of valuing into contact with Allan Gibbard's moral psychology, which postulates the state of norm acceptance, and the psychology I have proposed, which postulates the similar state of subscription to a standard. See Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990); and Copp, *Morality, Normativity, and Society.*

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

specific intentions and, when appropriate, form the intention to act, other things being equal. In this way, the account of valuing that I recommend makes transparent the relation between valuing, deliberation, and action. This is a theoretical advantage of the account.

In this section, I have been defending my view that our values are central to the autonomous regulation of behavior. First, I explained that our values are a kind of policy. Second, I argued that because they are intrinsic policies, to the extent that we are self-governed they govern our formation of specific instrumental plans that then constrain and guide our actions. Moreover, given this, and because of the way our values are components of our identity, cases in which we govern our actions by our values qualify as cases of self-government.

VII. Rationality and Values

According to the autonomy conception, the concept of rationality is closely related to the concept of autonomy. Rationality is a matter of doing well at self-government (understood in the thick sense) and at securing the requirements of self-government. Controversy would center on what governing oneself well consists in. I have been arguing that governing oneself (understood in the thick sense) is basically a matter of living in accord with one's values. If so, governing oneself well must consist basically in doing well at living in accord with one's values (and securing the requirements of doing so). This then is what I propose rationality to consist in. Ignoring certain qualifications, I propose the following standard, which I call the "values standard":

> A person is to serve her values as well as she can, overall and in aggregate, and in situations where more than one alternative would maximally contribute to serving her values, she is to serve her (other) intrinsic goals as well as she can, overall and in aggregate.[56]

To a first approximation, I hold that this is the standard that agents comply with insofar as they are rational—assuming that they have approximately accurate beliefs about what they value and about how to achieve what they value. It suggests the following principle:

[56] A person's goal is intrinsic if the person does not have it merely because she thinks that achieving it will be or may be instrumental to achieving other things that she wants. The formulation in the text ignores some of the caveats I mention in what follows. I need to leave the notion of "serving" one's values unexplicated. But see note 4 above. Important complexities lurk under the surface here. Philip Pettit has proposed that "serving" something one values might consist in "promoting" it or in "honoring" or "expressing" it. In some cases, it may seem more appropriate to "honor" something we value than to "promote" it. See Philip Pettit, "Consequentialism," in Peter Singer, ed., *A Companion to Ethics* (Oxford: Blackwell, 1991), 230–40.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

> Agent S is rationally required to do A in circumstances C just in case either (a) doing A in C is the action that would best serve S's values, overall and in aggregate, or (b) doing A is one of a group G of actions open to S in C, each of which would serve her values, overall and in aggregate, better than anything else she could do, and, of all the actions in G, doing A in C would contribute most to serving her (other) intrinsic goals, overall and in aggregate.

One might have various objections to this proposal, including the objection that a person's values can be pernicious. I will not be able to address every objection, but I will discuss those that seem most important. Before doing so, however, I need to mention four caveats.

First, there are constraints on the overall structure of a person's values, goals, and beliefs. To the extent that a person is rational, her values and beliefs are coherent with one another, and her (other) intrinsic goals are coherent with her values and constrained by them and by her beliefs, as well as coherent among themselves. Whether an agent is rationally required to attempt to achieve coherence in a given case depends on the costs and benefits of doing so, given the time and effort that would be required and given the agent's values and situation.[57] Second, a rational person serves her values as well as she can in light of the information available to her. She can be led by misinformation to believe that some action will serve her values when it will not, or to think that something will not serve her values when it will. Yet when her beliefs are reasonable, she may be acting rationally in acting on those beliefs.[58] To be sure, a rational person seeks information that is relevant to her capacity to serve her values and goals. She seeks such information to the extent that she is reasonable to believe that she should, given her values, her epistemic standards, and her epistemic situation.[59] And she assesses what to believe in a given epistemic situation in light of her epistemic standards or values.[60] Third, a rational

[57] There are complexities here that I must pass over. Compare John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 143. See Scanlon, *What We Owe to Each Other,* 121–22.

[58] Hence, the autonomy conception makes room for both kinds of error that Gavin Lawrence thinks any "proper theory" would accommodate. See Lawrence, "The Rationality of Morality," 121.

[59] A person's "epistemic standards" are norms she accepts that pertain to the rational formation of belief. An example is the rule to believe something only if the evidence or supporting reasons on balance make its truth significantly more likely than its falsity. By a person's "epistemic situation"—with respect to a given proposition—I mean her situation as it bears on the reasonableness of believing the proposition, given her epistemic standards.

[60] I pass over the problem of grounding epistemic standards. It is plausible that epistemic standards are appropriately grounded when, roughly, conformity with them would serve us well, in the typical circumstances in which we human beings actually find ourselves, in attaining true beliefs and avoiding false beliefs. On the account of practical rationality I am proposing, it is possible, even if unlikely, that a person who qualifies as practically rational might have epistemic standards that are not grounded in this way. For her epistemic standards may be such that conformity with them serves *her* well in pursuing what she values

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

person assesses her values and goals in light of new information when she believes she has reason to reconsider them in light of the new information. Fourth, where a person's serving her values (or goals) would put at risk her ability to sustain herself as an autonomous agent, rationality permits her to choose to sustain herself as an autonomous agent, by seeing to it that she is able to meet her basic needs.[61]

Even if we set aside these caveats, however, there is more to rational agency than merely acting *in accord with* the values standard, since this could happen by chance. Rationality is a matter of *governing* oneself well, and how well a person is doing at governing herself depends on the way in which she makes her decisions and not merely on what she decides to do. Setting aside various complexities, the basic idea is that rational behavior is *guided* by one's *values.* Such guidance seems to require a person to have at least approximately accurate beliefs about the content of her values, and to act in the belief that her action "makes at least as much sense," given her values, as would anything else.[62] In some cases, an agent is guided by her values in an explicit way. She deliberates about what to do, beginning with her values, considering how best to effect them, and finally reaching a decision about what to do. In many cases, however, an agent acts without consciously going through any reasoning. If she is guided by her values, there is something she values such that she does what she does *because* she thinks that her action *makes sense* in light of that thing. But she normally would not think that the key property of the thing she values is just that she values it.[63] She need not be guided by the values standard as such.[64] Thus, ignoring certain complexities:

---

even though it is not the case that conformity with them would serve human beings well, in typical circumstances, in attaining truth and avoiding falsehood. See Philip Kitcher, "The Naturalists Return," *Philosophical Review* 101 (1992): 53–114; e.g., 63. Compare Rawls, *A Theory of Justice,* 397.

[61] Meeting one's needs is a precondition for an autonomous life, so this qualification fits within the autonomy conception of rationality. I am not certain how the qualification should be worded. It should perhaps be restricted to emergency situations. The idea would be that rationality permits securing one's needs in emergency situations, even if doing so is contrary to serving one's values. For example, even if a rescue worker values his work, he may be rational to balk at going into a burning building when the risk to his life is excessive. By an emergency situation, I mean a situation with two key characteristics. First, it is reasonable for the agent to believe that serving her values would undermine or put seriously at risk her ability to sustain her status as an autonomous agent, not in a merely temporary or minor way, such that she could later compensate for neglecting her needs, but in a permanent and decisive way. And second, it is reasonable for her to believe that sustaining her status as autonomous would permit her to carry on in the future to serve at least some of her values. A version of the values standard that is qualified in the relevant way could be called the "needs and values standard."

[62] The locution "makes sense" is used in Gibbard, *Wise Choices, Apt Feelings.*

[63] For a related point, see Philip Pettit and Michael Smith, "Backgrounding Desire," *Philosophical Review* 99 (1990): 565–92.

[64] That is, a rational agent need not use the values standard as a decision procedure.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

> In a circumstance where agent S is rationally required to do A, and where S has approximately accurate beliefs about what she values, and about how to serve or achieve what she values, S is fully rational in acting only if (1) S does A intentionally, and (2), where V is something S values, S does A for the reason that, among other things, S judges that doing A will best serve V, and (3), where V, $V_1, \ldots V_n$, is the set of things S values, S judges that doing A makes the most sense given V, $V_1, \ldots V_n$.[65]

The reasons *for which* she acts should correspond to the self-grounded reasons that there are.

This suggests the following picture: A rational person—a person who complies with the values standard—governs herself on the basis of her values, making decisions that serve her values well given the priority she assigns to different values, and given the information she has. In situations in which all of her options would do equally well at serving her values, her values give her discretion, and she can do what will best satisfy her other intrinsic goals. She seeks information she needs in order to govern herself well, and she assesses the information and decides what to believe in light of her epistemic standards. She also seeks to sustain her ability to govern herself in this manner, at least insofar as doing so is compatible with her values, given her information. Rationality permits her to reconsider what to value, even if it calls for doing this in light of other things that she values. If a person values a life spent in the great outdoors, say, then, other things being equal, she lives this way despite the temptations of the city. But if she has values that would better be served by an urban life than by life in the woods, she might face difficult decisions. She might decide that her other values argue against the outdoors life, and she might give it up. Eventually she might cease to value it. This kind of change of values, in the interest of furthering one's ability to serve one's values, can be entirely rational.[66]

[65] I ignore the possibility of situations in which the agent judges that several actions would serve her values equally well. Where agent S is rationally required to do A because doing so will best serve her values, our assessment of the rationality of S's performance should be sensitive to, among other things, (1) whether S does A intentionally, (2) whether S does A for the reason that it will best serve her values, and (3) whether S believes that doing A makes the most sense given her values. In ordinary *akrasia,* the agent satisfies (3) but fails to do A. She may succeed, however, in serving something else that she values or takes as an end. Some cases of this kind will qualify as cases of "cleverness," as Wallace uses the term in "Normativity, Commitment, and Instrumental Reason," 1. In inverse *akrasia*, the agent succeeds on count (1), and perhaps also on (2), but believes that something other than A makes the most sense. In another kind of case, the agent succeeds on counts (1) and (3) but acts for the wrong reason. See Nomy Arpaly, "On Acting Rationally against One's Best Judgment," *Ethics* 110 (2000): 488–513.

[66] A similar point is made by Hubin, "The Groundless Normativity of Instrumental Rationality." David Schmidtz points out that having certain "final ends" can be a means to satisfying certain ends regarding our ends. See David Schmidtz, *Rational Choice and Moral Agency* (Princeton, NJ: Princeton University Press, 1995), 61, 77.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

VIII. The Objection from Ungrounded Ends

One might object to the values standard on the ground that a requirement to serve one's values can be plausible only for values that are themselves supported by reasons. People can have irrational and immoral values. Stephen Darwall writes: "If one had no reason to adopt A (or worse, reason not to do so), then maybe [instead of *serving* A], one should give up A." He writes: "From the facts that one has adopted A as an end and that B is a necessary means to A, it does not follow that one ought or has reason to B." It follows at most that one is required *either* to take the necessary means to A *or* to give up A.[67] Call this the "objection from ungrounded ends." In effect, the objection is that because people can have irrational and immoral values, compliance with the values standard cannot be a requirement of rationality.

I agree, of course, that the values standard needs to be grounded, and I shall propose a strategy for grounding it. This proposal will be the heart of my reply to the objection. There are, however, some preliminary responses to consider, the most important of which turns on a distinction between rationality and what I will call "wisdom."

It will be useful to begin by considering a related objection, the objection that the values standard needs to be amended to take account of *irrational* values. I agree that a *set* of values can be irrational—the values of a rational person must be coherent with one another. But I deny that a value can be irrational, just as such, merely in light of its content.[68] One might object that since I hold that the values standard is grounded in autonomy (in a way I shall explain), I ought to admit that counter-autonomous values—values that are not compatible with one's autonomy—are irrational. Perhaps it is irrational to have a policy of always deferring to one's parents.[69]

I do not want to accept this proposal. A person who has a policy of always deferring to her parents may be doing well at governing herself in light of this policy. I do not want to say that such a person is automatically irrational. For I want to distinguish objections to her values from criticisms of her success in governing her life in accord with her values. I say that objections to a person's values do not speak against her rationality. I do not want to view people from cultures that do not value autonomy as

[67] Stephen Darwall, *Welfare and Rational Care* (Princeton, NJ: Princeton University Press, 2002), 7. See also John Broome, "Normative Requirements," *Ratio* [new series] 12 (1999): 398–419; Patricia Greenspan, "Conditional Obligations and Hypothetical Imperatives," *Journal of Philosophy* 72 (1975): 259–76; and Korsgaard, "The Normativity of Instrumental Reason," 252. Korsgaard writes: "Hypothetical imperatives cannot exist without categorical ones, or anyway without principles which direct us to the pursuit of certain ends, or anyway without *something* which gives normative status to our ends" (ibid., 250).

[68] I ignore values that are logically impossible to satisfy. For useful discussion, see Donald C. Hubin, "Irrational Desires," *Philosophical Studies* 62 (1991): 23–44, esp. 24–26.

[69] I owe this objection to Steven Davis and Melinda Ammann.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

irrational simply on the ground that they have counter-autonomous values. To do so would blur the important distinction between rationality, understood as a matter of how one governs oneself, and "wisdom," understood as a matter of having values that are morally and otherwise acceptable. The objection from ungrounded ends also blurs this distinction, since it denies the rationality of serving values that are not based in reasons.

The intuitive basis of the distinction between rationality and wisdom is suggested by the Everest example. The climbers were not *wise* to give such importance to climbing Everest, but, despite this, we see them as having been rational in their pursuit of their goal. The values of a rational person may be subject to a variety of criticisms; they may be immoral or impolitic or self-aggrandizing or foolish. A good navigator may not be navigating toward a good destination. Of course, nothing turns on my choice of words to mark the distinction. It would not be a misuse of English to call the climbers irrational for giving such importance to climbing Everest, or to say it would be irrational to have a policy of deferring to one's parents. The important point is that there is a distinction between evaluating how well a person does at governing herself, given her values, and evaluating a person's values. I am concerned with evaluations of the former kind.[70]

Let me return, then, to the objection from ungrounded ends. I agree with Darwall that the fact that one has adopted an end does not entail that one has a reason to serve it, and I agree with Wallace that there is "no genuine requirement to take the means that are necessary for realizing ends that one merely happens to desire."[71] First, the values principle is concerned with serving values, not ends. A value is not an end that one merely happens to desire. A person might be rationally required *not* to pursue an end, if doing so would conflict with serving her values. Thus, the fact that I have a given *end* does not mean that I have any reason to serve it. Second, I can agree that it does not *follow* from the fact that I value something that I have reason to serve it. The values principle is a substantive principle that needs to be grounded.

[70] Gavin Lawrence discusses a view he calls the "end-relative account," which is similar to the conception of self-grounded reason that I am defending. He says that on this account "there is simply no question of whether an end is good or worth pursuing." See Lawrence, "The Rationality of Morality," 115, 128. I am claiming that it *is* possible to evaluate a person's values but that this is a different matter from evaluating the rationality of her actions. Korsgaard says that "the instrumental principle," which is similar to the values standard, cannot stand without "*something* which gives normative status to our ends." Yet she concedes that this status may be very thin. She allows that a "heroic existentialist" might endorse his ends for no "further reason," without thinking his ends are good. See Korsgaard, "The Normativity of Instrumental Reason," 251 n. 74, and 252. I have said that an agent "endorses" her values and the role of her values in governing her actions. This is not to say, however, that an evaluation of the rationality of an agent's actions in terms of the values standard requires or involves an evaluation of the agent's values.

[71] Wallace, "Normativity, Commitment, and Instrumental Reason," 1.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

To be on target, then, the objection needs to be reformulated. The objection is that I have not explained why or how it can be that "ungrounded" *values* are a source of reasons. Michael Bratman has pointed out that this objection may be especially difficult for me, given my view that values are general intentions or policies.[72] *Specific* intentions do not seem to be a source of *basic* or *underived* reasons.[73] If I form the intention to go to the gym, this gives me reason to take my gym clothes with me, but it does not give me a new reason to go to the gym. General intentions may seem to be similar: we adopt general intentions on the basis of reasons, but perhaps they are not sources of new reasons. The objection is that if values are just a kind of general intention, I need to explain why or how they can be a source of basic reasons when other intentions are not.

A fundamental response to these objections will have to wait until I discuss the grounding of the values standard. But I want here to bring out the plausibility of the idea that values can be a source of basic reasons—even if values are not themselves grounded in reasons, even if they are simply a kind of intention, and even if specific intentions are not a source of basic reasons.

Nevertheless, I think that a specific intention *can* be a source of a basic reason. If, on a whim, I form the intention to smell a nearby rose, this ordinarily would give me a reason to smell it, assuming that doing so would not conflict with my values. This reason would be basic, in that it would not be derived from any other reason.[74] But suppose I form an intention to act in a way that, I realize, conflicts with doing what would best serve my values. Suppose, for example, that I decide on a whim not to wear my seatbelt. Since I value my safety, I have reason *to* wear the seatbelt, and—in the given example—*no* reason *not* to wear it. Intuitively, even if I have no *reason* to value my safety, my value gives me a reason to wear the seatbelt, but my whim gives me *no* reason *not* to wear the seatbelt. A decision or intention motivated by a whim does not give a person a reason to act contrary to her values. Two factors seem crucial to explaining this.

[72] Bratman pressed this objection in personal correspondence.

[73] Scanlon, *What We Owe to Each Other*, 70. Bratman stresses that once one has formed an intention to A, one is under certain rational constraints with respect to this intention, such as the constraint not to change the intention without reason. But this is not to say that one acquires a new reason to A merely in virtue of forming the intention. See Bratman, *Intentions, Plans, and Practical Reason*.

[74] But would the reason be derived from a reason given to me by the whim or urge to smell the rose? There is a temptation to say so. However, an urge or whim can instead be viewed as an occurrence that is not in itself reason-giving. A person can decide whether to indulge it or not, and either way, arguably, she makes no mistake, other things being equal. If she forms the intention to indulge it, then, arguably, she thereby acquires a reason to do so. On this view, a person who stops to smell a rose on a whim may be no more responsive to reasons than would have been the case had she not stopped. Yet, of course, in the latter case, we might doubt her wisdom in eschewing an innocent opportunity for enjoyment.

First, values are intrinsic. Ordinary policies and intentions typically are formed because we see acting in accord with them as means to serving our values. This is why action in accord with our values is fundamental to governing ourselves in a way that action in accord with ordinary policies (and whims) is not. Second, and more important, our values are aspects of our identities in a way that our ordinary intentions, policies, ends, and whims are not. Failing to live in accord with our values negatively affects our self-esteem. In this sense, a person's values are linked to her sense of self. The self-esteem of a person who sees herself as having failed to comply with her own values is shaken to some degree by a sense of failure and disappointment in herself or a sense of shame or regret. A knowing failure to serve one's values is a kind of self-betrayal.[75]

Let me return, then, to the original objection, reformulated as the objection from ungrounded values. The objection is that a person who values A without having any reason to value A is rationally required, at most, *either* to serve A *or* to give it up; she is not rationally required to serve A. But consider the alternatives. A person does not have the *option* of giving up a value in the way that she has the option of giving up an ordinary goal, such as to smell a rose. We cannot change our identities at will, and our values are embedded in our identities. Of course, our values can *change*—even as a result of decisions we make. A person might rationally try to change her values in light of other things that she values. But we cannot *decide* to give up a value, or even an ordinary end, in the way we can decide what to do. At the point of action, our values are set. Hence, if a person with a value is rationally required either to give up the value or to serve it, then since she lacks the option of giving it up, it is plausible that she is required to serve it. It is plausible that to the extent that she is rational, she will govern her action in accord with it.

The point here is not simply that we cannot change our values at will. The reason we cannot change our values at will is, in part, that they are embedded in our identities: they are grounds of our self-esteem and aspects of our self-conception. Otherwise they would not function as they do, as compasses in our lives.[76]

IX. Grounding a Conception of Rationality

A theory of rationality can be viewed as proposing a standard or a norm, such as the values standard. The theory then claims that its pro-

[75] Similar suggestions have been made before. Donald Hubin has suggested an account of reasons for action based on what a person "intrinsically values," and he suggests that our values are "expressions of our selves." See Donald C. Hubin, "Hypothetical Motivation," *Noûs* 30 (1996): 47. Korsgaard has said that an adequate account of normativity "must appeal, in a deep way, to our sense of who we are, to our sense of identity." Korsgaard, *The Sources of Normativity*, 17–18. For Korsgaard, your identity is "a description under which you value yourself." See ibid., 101.

[76] See the discussion of giving up an ideal in Sarah Buss, "The Irrationality of Unhappiness and the Paradox of Despair," *Journal of Philosophy* 101 (2004): 186–87.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

posed standard *is* the standard of rationality—the standard that rational agents would comply with to the extent that they can, given their knowledge, just in virtue of being rational. This claim needs to be substantiated by providing the standard with an appropriate grounding. Let me explain what would be involved in doing this.

We can formulate any number of purely arbitrary standards. There is, for example, a standard calling on everyone to do a pirouette every night at midnight. We do not think that this standard corresponds to a normative requirement or that it has any bearing on how we are to act. To ground the values standard, I need to show that it is not similarly arbitrary. I need to show that it has a status in virtue of which it is relevantly "authoritative" or "normative."

To show this, I need to show, at least, that the values standard has a status such that those who fail to comply with it have thereby failed in a significant way. There is a problem specifying the *kind* of significance this failure would have. There are irrational people, so we cannot insist that it must be a kind of failure that *anyone* would be motivated to avoid. If the values standard is the standard of rationality, then a knowing failure to serve one's values would be *irrational,* but to say this would be question-begging and unhelpful. Given the content of the values standard, those who fail to comply with it face a loss, and perhaps a significant loss, since they fail to serve their values as well as they could. This is a kind of loss virtually anyone would be motivated to avoid, but would they be motivated in the way, or for the reason, we are motivated to avoid irrationality?

Donald Hubin suggests that a norm calling on us to pursue our values stands in no need of grounding.[77] He compares his view of the status of this norm to H. L. A. Hart's view of the status of the fundamental "rule of recognition" in a legal system. He seems to think that just as, for Hart, it is enough that the rule of recognition be treated a certain way by officials in the legal system, so it is enough that the norm calling on us to pursue our values be treated a certain way by us. At root there are simply certain brute facts, such as that we pursue what we value, and we at least sometimes assess our actions in terms of our values. And Hubin writes: "[T]he property of being rationally advisable just is the property of being properly related to these brute facts."[78] However, if the pursuit of what one values is "rationally advisable," then if a person fails to pursue what she values, this is not simply a departure from a brute psychological regularity. It is a failure in some interesting normative sense. The challenge is to explain this.

There is an analogous issue about politeness. Consider a rule that calls on us not to wear hats indoors. We might think that this rule is pointless

[77] Hubin, "The Groundless Normativity of Instrumental Rationality."

[78] Ibid., 467; see 463–68. See H. L. A. Hart, *The Concept of Law* (New York: Oxford University Press, 1961).

and deny that it has any bearing on how to act, even if we understand that compliance with this rule is locally taken to be a matter of politeness. Nevertheless, in a culture where there are rules that are taken to define politeness, there will be a widely accepted second-order standard, which we could call the "standard of politeness," that calls on people to comply with the local standards of conventionally acceptable behavior, such as the rule about hats. We cannot suppose that the standard of politeness has a bearing on how we are to act if we think that it is just as arbitrary and pointless as the rule about midnight pirouettes. Those of us who think that politeness has a bearing on how to act—that it is a significant failure of some kind to be impolite—must therefore think that the standard of politeness has some relevant authority or status.

There is an obvious way to think about this. The point of etiquette is to contribute to comfortable and pleasing social interaction. The standard of politeness is relevantly authoritative in virtue of the fact that compliance with it helps to make for comfortable and pleasing social interaction. It plainly is not the case that everyone values comfortable and pleasing social interaction, but the proposal is not that the standard of politeness is authoritative in virtue of the fact that we *value* comfortable and pleasing social interaction. The proposal is instead that there is reason to comply with the standard of politeness in virtue of its status as facilitating social interaction. Call this the "social interaction theory" of politeness.

A complete theory of rationality must be supplemented by a theory that has a similar form and a similar purpose to the social interaction theory of politeness. On anyone's view, rationality requires acting in certain ways, ways that could be expressed in a standard calling for us to do such and such in such and such circumstances. Any proposal about the content of the standard of rationality is incomplete unless it is accompanied by some account of the basis or authority of the standard.[79] This, then, is the grounding problem. To solve it, I think we need to understand the *point* of evaluating people and their actions as rational or irrational.

It might seem that if we can provide an analysis of the *concept* of rationality, and show that a given standard best captures our concept, we have done all that can be done to ground it. I hope that we can do something different from this. Suppose that the concept of rationality is the concept of the kind of virtuosity in the pursuit of what one values specified by the values standard. If so, then we rightly call people who comply with the standard "rational." But this does not yet give us any reason to view the standard as authoritative. Even if we rightly consider

[79] Elsewhere, I have proposed a schema according to which a basic normative proposition of kind K is true only if a corresponding K standard or norm has a relevant kind of status or justification. See Copp, *Morality, Normativity, and Society,* chap. 2. An account of the grounding of the standard of rationality could be plugged into this schema to yield an account of the truth conditions of claims about rationality. For further discussion of this idea, see Copp, "Moral Naturalism and Three Grades of Normativity."

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

ourselves "rational" insofar as we conform to the standard, we might wonder whether the standard has any status in virtue of which it actually imposes a requirement on us. Perhaps it will be replied that our concept of rationality is the concept of an *authoritative* standard—so that if the values standard *is* the standard of rationality, it *follows* that it is authoritative. But if this is correct, then if we are in doubt as to whether the values standard is authoritative, we are committed to being equally in doubt as to whether it is the standard of rationality.[80]

The issue is one that Kant apparently had in mind in the *Groundwork of the Metaphysics of Morals,* where he discussed the "possibility" of an "imperative." In my terms, a Kantian imperative is an "authoritative" standard, one that sets out an actual requirement. The possibility problem for Kant is to explain how to conceive of "the necessitation of the will expressed by an imperative in setting a task" and to show that there is such necessitation.[81] In my terms, that is, regarding a proposed standard of rationality, the problem is to explain how such a standard can be "authoritative" or "normative," such that it actually imposes a requirement.

X. Grounding the Values Standard in Autonomy

I have been arguing that the concept of rationality is closely related to the concept of autonomy. To be autonomous is to be *self*-governing, and to be rational is to do *well* in *governing* oneself. From one side, this suggests that the *content* of the standard of rationality should be determined by investigating what is involved in governing oneself. From the other side, it suggests a strategy for explaining the *grounding* of the standard of rationality.

The idea in outline is that rationality serves self-government—understood in the thick sense in which self-government involves regulating our actions by our values. The point of assessing the rationality of a person's actions is to appraise her success in manifesting or securing her self-government. That is, if the values standard is the standard of rationality, its grounding or warrant is that to comply with it furthers or instantiates one's self-government. Irrationality is a failure of self-government, a departure from governing oneself well. (Of course, irratio-

[80] Kant writes: "Whoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his [ends] and that lie in his power. This proposition, as far as willing is concerned, is analytic." Kant, *Groundwork of the Metaphysics of Morals,* 27 (Ak 417). However, unless there is a requirement of some sort that we be fully rational, then even if it is analytic that fully rational agents do intend the means to their ends, it does not follow that there is a requirement to intend the means to our ends. Hence, in Kant's terms, it does not follow that the standard calling on us to intend the means to our ends is an *imperative.*

[81] Kant, *Groundwork of the Metaphysics of Morals,* 27 (Ak 417). I discuss some of Kant's arguments in David Copp, "The 'Possibility' of a Categorical Imperative: Kant's Groundwork, Part III," *Philosophical Perspectives* 6 (1992): 261–84.

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

nal people may be autonomous in the thinner sense; they do poorly at governing themselves in the thicker sense.) The proposal is not that we are rational to manifest and secure our autonomy because we value autonomy. It is that we are rational to manifest and secure our autonomy, even if we do not value autonomy, because manifesting and securing one's autonomy is what rationality consists in.

Call the condition that a standard meets when complying with it furthers or instantiates one's self-government, the "autonomy condition." If what I have argued is correct, it should be plain that the values standard meets this condition. It calls on us to govern ourselves on the basis of our values, or, where our values do not dictate what to do, on the basis of our other intrinsic goals. To act in a way that would not serve our values as well as something else we could do would be to fail to control our actions on the basis of our values. But our values are policies we have that are central to our identity and that are intrinsic in that they are not grounded in any other policies or intentions. They are aspects of our self-conception. Thus, to fail to govern our actions on the basis of our values would be to fail in *self*-government. And in a case where more than one thing we could do would serve our values equally well, to fail to serve our other intrinsic ends as well as we could, while serving our values as well as we can, would be to fail to follow our own policies, despite being able to. This too would be to fail in self-government. Hence, it is plausible that compliance with the values standard furthers or instantiates *governing* oneself.

The autonomy condition is meant to play the same kind of role in grounding the values standard as the condition of furthering pleasing and comfortable social interaction played in grounding the standard of politeness. One might object that to establish genuine normativity, something much more ambitious than this must be done in order to ground the values standard. For if the grounding of the values standard is analogous to the grounding I proposed for the standard of politeness, then rationality is not a practical virtue that is superior to politeness. There are several things to say in reply to this objection, but I limit myself to two. First, recall that I am working here on the assumption that reasons pluralism is true. I think that there are reasons of etiquette just as there are self-grounded reasons. For this reason, I think it is an advantage that my grounding of the values standard is analogous to the grounding of the standard of politeness. Second, although I accept reasons pluralism, I agree that rationality is a special kind of practical virtue. It is special, not *metaphysically,* but because of its role in *deliberation,* as I will explain.

One might object that since many people and many cultures do not value autonomy, the autonomy condition cannot do the work I want it to do in explaining the authority of the standard of reason. For, one might think, a plausible account of rationality must be culturally neutral and neutral among our substantive values. Moreover, if some of us do not value autonomy, then the fact that compliance with my proposed stan-

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

dard of rationality would serve our self-government would not show that every person has reason to care about being rational.

In response, I want to say, first, that my account does not imply that it is irrational to fail to value autonomy. In fact, it implies (although with a qualification I am ignoring) that it is irrational to fail to promote satisfaction of one's values, *whatever* they are, even if one values a non-autonomous way of life.[82] According to the values standard, no one is ever rationally required to promote her autonomy unless her own values require her to do so. Second, no plausible theory of rationality can avoid implying that a person might be rationally required to conform with a standard she does not accept. For *any* theory needs to leave room for irrational persons. Finally, virtually anyone *would* want to do what she is required to do, according to the values standard. For the standard merely calls on us to serve our values. When we value something, we naturally are motivated to act appropriately, barring fatigue or depression or the like.[83] Motivation by our values is psychologically deep, for we tend to feel shame or guilt or disappointment in ourselves if we fail to act according to our values. Thus, to understand how we can be motivated to act in accord with the values standard, it is not necessary to suppose that we value promoting our own autonomy in addition to valuing such things as friendship. If irrationality is the failure of self-government, we can see why people typically care to avoid irrationality.

The objection seems to assume that the autonomy condition seeks to justify rationality as instrumental to autonomy. But this is a mistake. The view is, rather, that in being rational we *instantiate* our autonomy. The view is not that self-grounded reasons bind us only insofar as we *desire* to be autonomous, or only insofar as we *value* autonomy.[84] Many of us do not value being autonomous and have no desire to be autonomous.[85] Nor does my view depend on the idea that autonomy is *valuable*—though I do not deny that it is. To a first approximation, rationality requires us to serve our *values* because serving our values instantiates being autonomous, and because rationality consists in instantiating autonomy. This is

[82] I refer to the qualification about meeting one's needs. See note 61 above.

[83] Valuing something involves having a policy, or a general intention. One who has such an intention has at least a background standing disposition to act appropriately.

[84] If this were my view, then the argument would show, at best, that the requirement to comply with the values standard is hypothetical, or conditional on our valuing autonomy. This is not so, however. When we are rationally required to do something A, in most cases the requirement is conditional on, roughly, our having values such that doing A best serves those values in the circumstances. But the requirement to comply with the values standard itself is not conditional on our having any values in particular. Whatever our values, in complying with the values standard, we best serve those values.

[85] David Velleman argues to the contrary that agency has a "constitutive goal" such that any agent must aim to be in conscious control of her behavior—which he says amounts to the aim to be autonomous. It is in virtue of this fact that agents are subject to reasons. See J. David Velleman, "The Possibility of Practical Reason," *Ethics* 106 (1996): 719. See note 14 above. My account does not require postulating a "constitutive aim" of agency.

why, *whatever* a person values, she is rationally required to serve her values. My claim is not about what rationality requires her to value. It is about what rationality requires her to do *given* what she values.

XI. Deliberative Priority

My account assigns priority to rationality by comparison with other practical virtues, but not by assigning it a special *metaphysical* status. Instead, it assigns rationality a priority in *deliberation.* When agents with the necessary self-understanding and necessary information deliberate about what to do, the "default" is that their decision is in accord with the values standard.[86] Of course, in my view an agent who decides otherwise is to that extent less than fully rational. But the priority-in-deliberation thesis is that it is a law-like truth that, given the nature of practical deliberation, when agents who have approximately accurate beliefs about what they value, and about how to achieve what they value, deliberate about what to do and reach a decision based on their deliberation, other things being equal, they decide to do what they are required to do according to the values standard.[87] This is the default case.

This thesis is supported by claims I have made about what is involved in valuing something, given that the values standard calls on us to serve our actual values. First, values are policies or general intentions. We would not count you as valuing safety, for example, unless you had a tendency to act with caution when faced with known dangers. Second, your values are an aspect of your identity. This means that emotions of self-esteem are harnessed to your values and help to ensure that you have a tendency to pursue what you value. Moreover, the thesis is supported by a picture of the nature of practical deliberation—deliberation that leads to decisions about what to do. According to this picture, in central cases, practical reasoning involves reasoning from general intentions or policies to specific intentions in action. It is means-end reasoning in the sense that it concerns how to carry out general intentions. Given this picture, if a person is rationally required to A according to the values standard, then, other things being

[86] Korsgaard holds the stronger view that compliance with the "instrumental principle" is *constitutive of deliberate action.* She writes: "[I]f you don't put one foot in front of the other you will not be walking and you will get nowhere. . . . The instrumental principle is, in this way, a constitutive norm of willing, of deliberate action. If you are going to act at all, then you must conform to it" (Korsgaard, "The Normativity of Instrumental Reason," 249). But Korsgaard insists that failures to conform to the instrumental principle must be possible if the principle is normative (ibid., 247–48, 228). Surely, however, a person who fails to conform to the instrumental principle may have acted. Agents sometimes perform *akratic* actions. If the climbers had failed to serve their values in the Everest example, but had instead stopped to help the victims, their helping would have been an action. It therefore seems implausible that conformity with the instrumental principle is constitutive of "deliberate action."

[87] The idea of a defeasible law-like generalization is discussed in Mark Lance and Margaret Little, "Defeasibility and the Normative Grasp of Content," *Erkenntnis* (forthcoming).

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

equal, she would form the intention to A if she were to deliberate cogently about what to do in light of her values.

There can be exceptions. The person might be depressed or exhausted and might for this reason fail to form any intention to act. Or she might fail to reach any conclusion about how best to serve her values. She might be unable to see how to resolve conflicts among her values. She might not understand what she values, or she might lack relevant information about how to serve her values. Deliberation might lead her to become perplexed about what she values. She might lack the fortitude to resist an impulse to act contrary to her values. She might be *akratic.* Or, perhaps paradoxically, she might *have* the strength to resist the impulse to act in *accord* with her values. In a moment of clarity, she might see that her values are morally unacceptable and decide to do the right thing. The climbers in the Everest example might decide to abort the climb and to help the victims, thereby acting out-of-character and irrationally, but doing the right thing.

With these caveats understood, we can see that the account of rational agency I have proposed explains how reasoning about what to do can lead to a decision, and thus to the forming of an intention. And we can see how rational decision-making can be the default case, assuming approximately true relevant beliefs. Our values are partly constituted by general intentions or policies, and the default is to be guided by them to form specific intentions that will implement or further or express our values. Suppose you are trying to decide whether to watch a tennis match or to read a novel. Your decision likely will turn on such considerations as how much you are enjoying the novel, how often you find time to watch tennis or to read, who is playing in the match, and so on. Let us suppose that you value the simple pleasures of life, which, for you, include reading novels and following tennis. If you realize that this is the only thing you value that will be affected by your choice, you will pay attention to which of the activities in question promises more enjoyment. Once you have reached a conclusion about this, if all goes well, your value—which is, inter alia, a general intention—will lead you to form a specific intention either to read or to watch tennis. In the default case, if you have been reasoning with accurate information, your intention will be to do what the values standard implies that you rationally ought to do.

The important point here is that self-grounded reasons—facts about the impact the agent's alternatives would have on the things she values—have a role in deliberation that other kinds of reasons do not generally have. When agents who have the necessary self-understanding and the necessary information decide to act, the default is that their decision is in accord with the values standard. The corresponding thing cannot be said of morality or etiquette, for example. It is not the case that when a person with the necessary self-understanding and information decides what to do, the default is that she decides to act morally or politely. This is true only of people with the corresponding values.

One might object that a person who is deliberating about what to do is trying to decide what to do *period,* not merely to decide what *rationally* to do. To see this, consider again the Everest example. Given the values of the climbers, suppose that the values standard would require them to press on to the summit. This would mean that carrying on is what the climbers ought *rationally* to do. But I have been supposing that they ought morally to stop and help the victims. One might object that a reflective climber would want to decide what to do *period,* not merely what she ought rationally to do. On my own view, it might seem, the values standard cannot tell the climbers what they ought to do *period.* For as I said earlier, I hold that there is no answer to the question of what the climbers ought *simpliciter* to do in the imagined situation. That is, I am assuming that there *is* an answer to the question of what they ought rationally to do, and that there *is* an answer to the question of what they ought morally to do, but I hold that there is no overarching normative standard that determines what they ought to do *simpliciter.* Now it is not part of my goal in this essay to defend my view about "ought *simpliciter,*" so I could avoid the objection by giving up my view. I do not need to do this, however; for the objection is based on a misunderstanding.

The upshot of practical deliberation is not a belief about what one ought to do; it is a decision or an intention. On my view, the values standard cannot tell the climbers what they ought to do *simpliciter,* but only what they ought *rationally* to do. But the climbers are trying to decide what to do. They are not trying to decide what to believe to be the rational thing to do. And a climber's decision about what to do would be a flat-out decision, a decision either to help the victims or to carry on with the climb. I agree that a reflective climber would want to decide what to do *period,* but on my account she would do so. That is the nature of decisions.

It is true that if someone decides to do something, and if her deliberation is relevantly informed and fully rational, then she ought *rationally* to do it. This does not entail, however, that she ought *simpliciter* to do it. For deliberation is carried out from one's own practical standpoint, not from the standpoint of a metaphysically overarching standard that determines what one ought to do *simpliciter.* I have argued, in effect, although with caveats, that one's practical standpoint is constituted by one's values. If a climber decides to carry on with the climb, she would be confused if, after the climb, she claimed to have made the mistake of failing to decide what to do *period.* Her decision may have been morally indefensible, but it was a flat-out decision. Her mistake was a moral mistake, not the mistake of failing to make a flat-out decision.

XII. Conclusion

I have distinguished two issues that a theory of rationality must address. First, it must specify the content of the standard of rationality. Second, it

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press

must ground the standard in order to support its claim that the standard is the standard of rationality. I have proposed the values standard, and I have claimed to ground the values standard in the autonomy condition. *Rationality* is in the service of self-*government*, I have argued.

In closing, I would like to emphasize the modesty of the view. According to the autonomy conception, rational agents are necessarily disposed to comply with self-grounded reasons, just in virtue of being rational. But rational agents are not necessarily disposed to comply with moral reasons, even if they recognize that there are moral reasons that bear on their actions. A rational person *could* have moral values, so moral reasons *could* be reasons for which she acts. A morally virtuous agent has exactly the moral values it is morally best to have, and thus, for such a person, moral reasons are also self-grounded reasons. It is not necessarily true, however, that rational agents have moral values, so it is not necessarily true that they are rationally required to do what they believe they have moral reason to do. A person who ignores reasons of this kind is not necessarily failing to govern herself in light of her values.

Philosophy, University of Florida

https://doi.org/10.1017/S0265052505052076 Published online by Cambridge University Press