Segmented Agency*

Elijah Millgram

Department of Philosophy
University of Utah
215 Central Campus Drive
Salt Lake City UT 84112

lije@philosophy.utah.edu

December 15, 2013

A great many arguments in contemporary analytic moral philosophy turn on agency, and appeals to one view or another of what an agent is are used to defend positions regarding, among other topics, practical reasoning and substantive moral theory. One indicator of how heavy the traffic has been is the inevitable backlash, with a debate emerging as to whether anything can really be accomplished by invoking agency at all.[1] Here I want to take a somewhat different, and perhaps more constructive tack. I am going to argue that the more-or-less shared conception of agency in play is mistaken, but the point of doing so will be to ultimately arrive at a corrected conception of agency, one which can be put to better philosophical use.

The mainstream view in moral psychology is that agents are unified, both in fact (for the most part) and ideally (disunity of agency is regarded as a defect). Accounts of agential unity vary, and include as components

*I'm grateful to Tom Blackburn, Luca Ferrero, Alisa Garcia, Karen Mottola, Tom Pink and Margaret Stock for helpful discussion, to Chrisoula Andreou, Michael Bratman, Teresa Burke, Sarah Buss, Christine Chwaszcza, Kelly Heuer, Jenann Ismael, Kimberly Johnston, Matt Mosdell, Stephanie Shiver and Ben Wolfson for comments on earlier drafts, to audiences at Tel Aviv University, the University of Cologne, the University of Arizona, the University of Bern, the University of the Saarland, the University of Utah's Kaffeeklatsch, to the participants of the Templeton Foundation Summer Retreat on Love and Human Agency, and, for reading through *Structures of Agency* with me, to my Fall 2008 seminar on Practical Reasoning.

[1] Millgram, 2005, Enoch, 2006, Ferrero, 2009, Tubert, 2010.

such theses as: agents produce actions which can be in a very robust sense attributed to them, actions they *own* because the agents are identified with, rather than alienated from, their choices; agents have 'practical identities' or 'ground projects' which they may lose, but cannot disown; agents do not pursue projects at cross-purposes with one another; having made a decision, they follow through on it (and do not instead act on some contrary impulse); they possess a unified point of view from which they render judgments about what is worth doing and what they will do; they reflect on their actions, and endorse their choices when they do; when they act, they act so as to understand what they are doing and why; their choices are governed by policies which dictate how competing reasons will be taken into account.[2] Across these variants, we find a rough but shared picture, of a creature that has integrated its goals, evaluative judgments and other guidelines into a single and internally consistent pattern, and whose control structures generate actions that are consistent with the pattern.

In order to avoid getting lost in the crosstalk, I am going to organize this discussion around just one of these views, a systematic account of agency developed in recent work by Michael Bratman.[3] There are a number of reasons for the choice. First of all, the materials that Bratman uses in his theoretical constructions (plans and policies, and I will explain just what these are shortly) are as clear and straightforward as one gets in this business, but also much more pliable, much less brittle material than competitors' devices.[4] Second, Bratman's treatment incorporates, reconstructs and criticizes other people's ideas—there's no Not Invented Here Syndrome, a rare virtue in philosophy—and engages other philosophers with opposing views (rather than, as so often happens, pretending they don't exist). Third, Bratman's style is consistently low-key and sane-sounding; this is particularly evident in his adaptations of other philosophers' ideas.[5] Consequently,

[2]See, for instance, Williams, 1981, Velleman, 2000, Korsgaard, 2009, Korsgaard, 2008, chs. 1, 3, Korsgaard, 1996a, ch. 13, Korsgaard, 1996b, Frankfurt, 1988.

[3]Bratman, 2006; freestanding page references will be to this book.

[4]For Bratman's materials, see Bratman, 2001, Bratman, 1987, and Bratman, 1999. To give you a sense of the contrasts I have in mind, Korsgaardian constitutions are an example of a less-straightforward theoretical construct (Korsgaard, 2009, Korsgaard, 2008, ch. 3), and Frankfurtian hierarchies of higher-order desires are an example of a clear but brittle construction material (Frankfurt, 1988, ch. 2).

[5]For instance, when Velleman goes looking for a mental state that can't be disavowed (2000)—and in a moment we'll gloss that phrase as well—he settles on a state and a matching theory that strikes everybody who encounters it as (sorry, David) *wacko*. When Bratman adopts the idea that we need to identify a psychological state, or complex of them, that isn't disavowed, he comes up with the sort of policies we will be taking up in a moment; the move seems thoughtful, plausible, and anything but wacko. (Notice the

Bratman's account of agency ends up being a best presentation of the central ideas in the current debate, and that means in turn that if we can see how they don't work out in his rendering of them, we will be well-placed to see how (as I am going to suggest) they *really* don't work out.

Unified agency has standardly been contrasted with fragmented agency, and that contrast has made unity seem nonoptional; if you're a highly fragmented agent, you have earned, as a comedian once put it, "the demeaning epithets that are said about people who are peeling an empty banana": the lights are on but nobody's home; you're not all there; you're not playing with a full deck; you're leading (and here is one from Bratman) a "seriously fractured life".[6] I will be introducing another and much more viable option: human agency is *segmented*.

I will begin by laying out Bratman's account of agency, and highlighting a number of its more important features. Then I will sketch a strategy, *serial hyperspecialization*, that human beings, as a species, adopt early and often, and indicate how it fits badly with Bratman's account. I will argue that, when it comes to what matters and what your reasons are, if you're ever *really* wrong, wrong all the way down, that's not—on the mainstream way of thinking about these things—something that you can properly face up to: that is, face up to by deciding on a new and different way of handling situations you are in. Finally, I will return to the idea that a serial hyperspecializer normally lives its life in segments: this means, I will suggest, that we are not only missing half of our theory of agency, but that we have been looking in the wrong place for a theory of personal identity.

1

Bratman's ambition is to use an uncontroversial and unproblematic philosophical toolkit to reconstruct and thus to explicate a handful of central notions and distinctions in recent moral philosophy. I'll first briefly rehearse his agenda; then I'll lay out the toolkit; then I'll describe the construction. Last, I'll redescribe the view of agency which Bratman exemplifies at the higher resolution which his reconstruction permits.

First, a good deal of effort has been devoted, over the last four decades, to making sense of, as we might call it, the superlative attribution of both

adjustment from Velleman's "can't be" to Bratman's "isn't", marked at 188 (esp. n. 5); however, as we will see in due course, there is a sense in which one cannot in fact disown the policies which Bratman is introducing.)

[6] P. 298; Martin, 1998.

actions and attitudes to an agent. You blurt out an offensive remark, but insist, later on, that you didn't really mean it; the attitude it expressed was *there*—as Bratman sometimes puts it, it was a wiggle in the psychic stew—and in some minimal sense it was your opinion: who *else's* would it have been? But it wasn't a view you endorse, and you can legitimately disown it. Some very strongly felt desires are for things that you are quite clear you don't really want: an apparent oxymoron, accommodated by the distinction between *mere* wanting, and *superlative* ('real') wanting. Some would-be beliefs and desires *are* full-fledgedly yours, and that you on reflection endorse them is evidently either strong evidence for the superlative attribution, or (perhaps in part) constitutes the attitude's "agential authority." Some actions are things you really did—as opposed to others which, although in some sense you performed them, just kind of happened, as you sheepishly tell others afterwards.[7] Bratman's first agenda item is to reconstruct this distinction, between 'really yours' and 'merely yours'—between attitudes and actions with which you identify, and those you can reasonably disavow.

Second, moral and political philosophers have been, again over about the last half-century, very much interested in autonomy.[8] So many philosophical demands have been made of this concept that it is perhaps unreasonable to hope to provide an account that covers everything that anyone has insisted must be part and parcel of autonomy, but self-government is, by anyone's lights, a central aspect of it, and, Bratman thinks, the aspect with which we should most concerned. So he hopes to provide a model of a self-governing agent.

Third, Bratman wants to be able to say what it is to value something. Partly the point of doing so is to make room for an observation about superlative attribution that we now think of as Watsonian: that identification is evaluative.[9] And partly (I am guessing) Bratman is tempted by the prospect of headway on longstanding open questions in metaethics.

In developing his account, Bratman adopts one widely shared constraint, and drops another; I need to pause to explain what these are. First, philosophers in this field experience pressures toward imagining people as ghosts who float above or stand behind themselves, and who intervene from time

[7] For an overview of some of the attempts to distinguish between full-fledged actions, and less-than-actions that count only as 'mere activity,' see Millgram, 2005.

[8] See, e.g., Hill, 1991, ch. 4, Dworkin, 1988, Christman, 1989; Fleischacker, 1999, describes its topic as 'liberty,' but can plausibly be taken as an account of autonomy as well.

[9] Watson, 2004.

to time in the workings of their own minds. For instance, in the literature triggered by Harry Frankfurt's early paper on personhood, what makes my desire superlatively mine is that it is endorsed by a second-order desire.[10] But what makes that second-order desire superlatively mine, rather than merely another wiggle in the psychic stew? A further desire? Won't *I* at some point have to decide that one of the desires in this regress *is* really mine? And when I admit to doing that, am I not representing myself as outside the natural order (taken here to include psychological states): something over and above, or concealed behind, the complex of psychological facts? In the face of this sort of pressure, Bratman is taking on board the requirement (which he sees to be correctly posed by Velleman) that his account not invoke, over and above the psychological structures that it adduces, the further presence of *the agent himself.*[11] The agent will turn out to be a psychological structure of a specified sort, one which will leave no such leftovers or loose ends.

Second, metaphysicians of agency typically insist that their theory of agency tells you what agency *is*, the implication being that there could be no properly so-called agents who do not conform to the theory. Bratman is attempting, more modestly, to provide *a* model of agency that exhibits features of interest, such as self-government. Although we are not shown alternatives, for all we know, there might be other psychic structures that amounted to, say, forms of autonomous agency supporting a distinction between really and merely wanting something.[12]

In earlier work, Bratman introduced plans as his preferred philosophical medium or rendering tool. *Plans* (equivalently, intentions) are stable controllers of conduct: that is, once you have adopted a plan, you need a special reason to reconsider and discard it, and if you don't reconsider, you can be expected to go ahead and do what it says. A *policy* is an open-ended plan, one which specifies that in cases answering to such and such a description, you will do so and so. *Self-governing* policies are higher-order policies; that is, they coordinate and manage lower-order, vanilla plans and policies, along with motivating attitudes such as desires and emotions, giving green lights to some of them and red lights to others. *Effective* policies bring it about that choices and deliberation comply with them: when a policy

[10]That is, a desire whose object is another desire: the idea is that I don't just want it, I *want* to want it—or, more carefully, I want to *act* on my wanting it. See Frankfurt, 1988, ch. 2.

[11]Pp. 24, 196; as he also puts it on occasion, the account must be "nonhomuncular" (p. 187; see also p. 177).

[12]See, e.g., pp. 163 (and esp. n. 5), 183, 197, 199.

is effective, the fit between policy and action is not merely fortuitous. A policy is *weight-bestowing* when it gives a prospective reason a weight in deliberation, and more generally, a higher-order policy is *reason-determining* when it determines that a first-order attitude (normally, a desire) play a specific role (defaultly, an end-setting role) in deliberation.[13] (An example, to get the general idea across:[14] a recruitment committee might decide that, in the course of deliberating about whom to hire, the department will give an applicant's area a great deal of weight, will treat affirmative action considerations as a tie-breaker, and will not treat applicants' rank as a consideration at all.) You are *transparent* when the cognitive functionality of your policies is captured in their content.[15] Your self-governing policies are *reflexive* when they endorse their own effectiveness.[16] (Again to give the general idea, if a department adopts the policy that procedural decisions require a consensus, reflexivity would require that that policy be supported by a consensus.) You are *satisfied* with a self-governing policy when it does not conflict with another of your self-governing policies. Lastly, an attitude, such as a desire or an intention or a policy, is *noninstrumental* if you have it

[13]Weight-bestowingness is adapted from Nozick, 1981, ch. 4; Bratman tends to use longer locutions to mark such policies (see, e.g., p. 142), and so the second term, adopted in the interest of terseness, is my own. Sometimes in Bratman's writing (as at p. 295), but not always, "self-governing" covers policies that are weight-bestowing and reason-determining.

Two observations: First, although Bratman does not emphasize it, these two concepts travel together; setting the weight of a prospective reason to zero is determining that it is not to play a motivationally effective role in deliberation, and conversely, giving it a positive weight is determining that it will. Second, although the discussion is conducted largely in terms of weights, restricting the policies in question to setting weights would be unnecessary and unprincipled. Just for instance, one could adopt the policy that a particular kind of reason is to be lexically ranked over another, even though lexical rankings are not representable by weights. The point is acknowledged at p. 300, but not systematically reflected in Bratman's terminology or illustrations.

[14]A variant of one introduced at p. 301.

[15]For transparency, see pp. 181, 191f. This condition excludes cases like this: you might have a policy of not starting to write until you've sharpened your pencils, made yourself a cup of coffee, read all the relevant literature, and reorganized your filing system. The real cognitive function of the policy is procrastination: it's a way of not starting to write anytime soon, and maybe never starting to write at all. But the policy doesn't say this anywhere, so to speak, and when you neurotically begin to sort through the filing cabinets, you may be unaware of what you're really doing.

Philosophers don't generally share an understanding of what it is to be conscious of something. So bear in mind that unless you have a matching theory of consciousness (the treatment of Korsgaard, 1996a, ch. 13, would perhaps suit), action governed by a Bratmanian policy need not be *conscious* action.

[16]See pp. 183f, 189f, 194, 211, 242.

not simply because it's expedient to have the attitude itself (as when someone is paying you to have it).[17] Just to have some shorthand for all this,

[17]See pp. 84, 190 (at the latter, Bratman is discussing the negative case, of inconvenient desires that give you reasons, not to fulfill them, but to get rid of them). The casual reader may be unaware how the requirement is an attempt to steer around a very large philosophical iceberg; here I will just take time out to explain what the iceberg is.

Let's allow for the sake of the argument that there are evaluations whose correctness is agent-independent, and where the direction of explanation runs from the correctness of the assessment to the agent's attitudes, rather than the other way around; call these (following Bratman) 'value judgments' (p. 172). Now, even if there are correct value judgments galore, it's implausible that, taken together with all the facts you have at your disposal, they settle many decisions you have to make: there's *slack*. When you drive up to San Francisco, you could take either 101 or 280, and the value judgments don't settle which it will be; you just have to form an intention, and follow through on it (e.g., pp. 159, 166, 205, 212, 233). But what goes for simple, garden-variety plans to drive up the Peninsula goes double for Bratmanian policies: it is very hard to believe that the value judgments on which you rely in choosing a Bratmanian policy uniquely determine which Bratmanian policy you choose. Bratman takes it for granted that our surplus-value commitments can do a great deal of work for us. The visible tip of the iceberg is that it seems obvious to many people that you can't adopt intentions arbitrarily—for instance, because someone is paying you merely to intend something (e.g., Kavka, 1983, Millgram, 1997, ch. 2).

The formal problem lurking below the water is this. On the one hand, when Bratmanian policies are stipulated to be noninstrumental, that condition is meant to rule out (roughly) cases in which the policy is not held for *proper* reasons, but rather because the policy is expedient to adopt. But on the other hand, policies take up slack, which is to say that when we act on their basis, we are acting not because we have *proper* reasons so to act, but because it was expedient to adopt some such policy, and the policy we adopted dictated acting in this way. How are we to square the need for the condition with the slack-assuming role of Bratmanian policies?

The substantive worry is that a Bratmanian policy won't really be *yours*—you won't be able to take it seriously—unless you think it really is a good idea, and a good idea largely *because* the things it says are reasons, really *are*. Accepting the policy in any other way puts one in a posture that philosophy professionals have no doubt at one time or another taken towards policies adopted in department meetings. ("Resolved that we will not consider the candidates' strengths in history while making this hiring decision.") One does it, but one doesn't really think these are one's reasons: because one is playacting, the policy doesn't speak for one (even when it speaks in one's name). Or—a slightly more elaborate example—think of polite fictions, such as the pretense required of faculty by the honor codes of some universities, to the effect that students are not cheating on their assignments. You have to pretend, in every way, that the student did his own work (while the Honor Board holds a 'trial'); you know that this procedure is the negotiated settlement of a war between the faculty and the students back in the 1930s, and you know it's settled; it's your job, and you pretend. But you're pretending: no matter how thoroughly this policy controls your inferences (or 'inferences'), no matter how unwilling you now are to change it, the 'belief' that the student isn't cheating isn't your belief; you're sure he *is* cheating! Bratman briefly considers what it's like to have this attitude towards a group decision in which one is participating (307f), but the important question here is what it's

let's call an effective, transparent, reflexive, noninstrumental, self-governing, weight-bestowing and reason-determining policy with which you are satisfied a *Bratmanian policy*.

With this notion in place, we can expeditiously review Bratman's reconstructions of the items on his agenda. You *identify* with an attitude (a desire in particular, but the account ought to generalize) if you have a Bratmanian policy to the effect that the attitude should play a designated role in your practical reasoning: in the case of a desire, that it sets ends for you, with such and such weight or force. You *value* something when you want it, and you have a Bratmanian policy in favor of treating your desire as a justifying consideration in your practical reasoning. Acting on the basis of deliberation governed by a Bratmanian policy is *self-governed* action, which is, once again, autonomous action in the sense most of interest to us.

The pivotal role of Bratmanian policies in these reconstructions of identification and so on is motivated by the thought that what makes you the same person over time is psychological continuity. (That is, Bratman is accepting what these days gets called a neo-Lockean approach to personal identity.) In Locke, it is in virtue of the memories you inherit from past selves that you are the person you once were; recent revivals of the view allow other states to count as well, and intentions in particular: that in the future you will act on decisions you now make contributes to that future self being *you*. Now, policies, as a variety of plan or intention, are stable, and so tend to persist across time. Self-governing policies, which regulate and control one's other, garden-variety plans and intentions, function as a sort of cross-temporal spine of one's agency: it is not just that you do something tomorrow because yesterday you decided to do it, but that you decided to do it because, long ago, you adopted a policy about what reasons were going to figure into your decisions. Because such policies largely constitute you as a person who persists over time, and because they are in the business of endorsing and disavowing, when they speak, they speak for you. So what it is for you to identify with a desire, say, or an action, is for a policy of the type we have just specified to endorse it.[18]

---

like when there are no other participants to distance yourself from: when it's *just you*. Surely a life lived on the basis of Bratmanian policies adopted in this sort of way is life according to *Dilbert*. But then, since the function of plans and policies, in Bratman's account, is to take up slack, aren't your Bratmanian policies normally to some extent or other adopted in this sort of way?

[18]Bratman qualifies the claim by allowing that other psychic structures—he calls them "quasi-policies," and mentions ideals as one possible type of quasi-policy—might serve an organizing role very similar to that of Bratmanian policies. We are not told much about the workings of these alternatives, but in sec. 5, I will pause to verify that quasi-policies

Here, then, is the picture of agency we are being offered (and, once again, we are interested in it because it is a best representative of a family of models of agency that amounts to today's philosophical common sense):[19] The agent is organized by and around a set of mutually compatible long-term policies. These policies specify what counts as a reason, and how much of one, when you're making up your mind what to do; they lead you to act on (or make you balk at acting on) other garden variety desires and intentions; they are reflexively self-endorsing. Because these policies are such important contributors to your personal identity over time, and because they are policies of endorsing or disavowing reasons for action, what it is for you to identify with an attitude—for it to be *really* rather than *merely* yours—is for it to be endorsed by such a policy. What it is for you to value something is for you to identify with your desire for it. And an agent who acts on the basis of attitudes that are superlatively his, in pursuit of things he values, is autonomous.

2

At this point, I want to pause to flag a handful of objections the reader is likely to be entertaining, both in order to forestall what would otherwise be ongoing sources of distraction, and to foreshadow moves and arguments that we will be taking up in due course.

First, recall the move from, A Bratmanian policy holds me together, and constitutes me as the same person over time, to, When one of my Bratmanian policies speaks, it speaks for me. (Call that Bratman's *Master Move*.) Bratman does not give an argument proper for his Master Move; it is usually marked with some such phrase as: "This suggests the conjecture. . . ," "This makes it natural to suppose. . . ," "Or so it seems to me reasonable to say." But evidently an argument is needed, because there are a great many things that are held together by components that do not thereby get to speak in their names. Books are held together by their covers, but what the cover says is not necessarily what the book says. States are held together by their police forces, and churchs, sometimes, by their censors, but what the police or the censors endorse is not necessarily what the state or the church endorses. Even when the police department announces the very

do not, so far as we understand them, short-circuit the argument.

[19] Though it is also part of philosophical commonsense to have qualms about how psychologically realistic such elaborate constructions can be, as when Watson remarks: "That most people have articulate 'conceptions of the good,' coherent life-plans, *systems* of ends, and so on, is of course something of a fiction" (2004, p. 25).

same thing as the legislature's designated spokesman, it is the spokesman who is speaking for the state, and not the police department.

Cases like these can serve as models for what looks like a direct counterexample to Bratman's proposal. Fantasies don't normally speak for one (although they can of course be very revealing), and one can have fantasy policies; for instance, I might while away the lazy summer afternoons working up self-governing policies that state what ends will have what weights in my deliberation, in circumstances that I'm quite convinced will never come to pass: for instance, policies that cover how I would deliberate if I suddenly had large amounts of money. ("Whenever I'm approached by a development officer, I'll take it to be a weighty reason to offer his institution a large gift that the naming opportunity involves an embarrassing title"—i.e., I'm daydreaming about endowing the Foolish Professorship of Philosophy.) Because I continue elaborating the fantasy policy, year after year, it is a major contributor to my neo-Lockean personal identity; in the sense at hand, this fantasy (and others like it) hold me together, and constitute me as the same person over time. I have no inclination to change the fantasy, and so I'm satisfied with the policy; moreover, in the almost unimaginable event that I did get the money, I would probably act on the policy, out of sheer inertia; after all, it's not like I have other, competing habits and plans in place for handling large amounts of money. Nonetheless, the psychic function of this policy is that of a *fantasy*, and I can be entirely aware of that; when it speaks, it doesn't speak for me.[20]

The more minimal version of the point is that the argument for Bratman's Master Move is missing, and we shouldn't buy into his position without one. But there is another way to take the problem. In order to determine whether a distinction is correctly tied to an account of personal identity, we would need first to be clear about what the distinction is *for*. Now, there is a shared assumption at work behind the various explications of superlative

[20] Philosophers these days seem insufficiently sensitive to the distinction between fantasy and other 'pro-attitudes', especially desire. The most important difference between them is that daydreams are themselves consumed, whereas desires direct one to an object that is to be consumed. For instance, I'm told that a recent Bond movie has free running (or parkour) scenes in it. (A free runner will run up, say, a crane, jump off the end onto the ledge of a building, run along the ledge. . . .) Presumably the scenes are there because people enjoy the fantasy; Bond movies are canned fantasies that people pay $10 or so to see. But I'm pretty sure that most of the audience members, if offered an opportunity to engage in free running, would decline an activity they are quite aware will be scary rather than enjoyable. They don't want to free-run; they just want to consume the fantasy, and more generally, people typically consume fantasies of things they don't desire. (I'm grateful to Elizabeth Calihan for bringing this to my attention.)

agency, namely, that the difference between mere attribution and superlative attribution is *important*: that it makes a great deal of difference whether an attitude—belief, desire or policy—speaks *for me*. Explaining the alleged link between superlative attribution and personal identity requires us to step back and ask why we *care* about superlative attribution.[21] This of course can be a rhetorical question,[22] but although I am sensitive to the ways that philosophizing on this topic sometimes looks like the product of an obsession, rather than well-motivated problem solving, I do not intend it that way. When the time comes, I will propose a function (not, I expect, the only one; I mean only to be making a start on the problem I'm posing) for the distinction between what you really want, believe and do, and what you in some lesser sense want, believe and do.

Turning now to a second and third problem: Recall that being 'satisfied' with your deliberation-governing policy is a requirement imposed on its speaking for you, on its giving rise to autonomous action and so on. The idea is that whether some candidate self-governing, temporally cross-referencing policy is really your own is only a live question if you've got a bad attitude about it—in this case, literally an attitude, in the philosopher's technical sense. After all, if you don't have the bad attitude, you can't plausibly claim to be alienated from your policy. In Bratman's appropriation of Frankfurt's

[21] Although Bratman himself does not explain why the distinction matters, other theorists do, and the point is usually forensic: we can only be held *responsible* for what we think, decide and do in the full-fledged sense. (We can name enterprises with this sort of orientation *Perp Theory*.) However, the fit between the sorts of hierarchy-oriented theories we find in this debate and the forensic social function is bad. The distinction, forensically construed, is important in the first place because it *has* a social function: if you do something rude, sometimes you get to say, "I'm sorry, it wasn't me speaking—it was the booze." When you disown your attitude or action, it can have real social consequences, and therefore this social function couldn't be served by a subtle and complicated distinction made out at the upper reaches of a hierarchy of attitudes in the privacy of one's mind. If those who invoke the distinction aren't going to be *getting away with stuff*, the legitimacy of the excuse has to be something that others can check on.

[22] As it is in Foucault, 1984, which reminds us that the concern for claiming authorship is a recent cultural phenomenon. (Recall that one prominent agency theorist, Christine Korsgaard, adopts the vocabulary of 'authorship' to mark superlative attribution: you *author* your actions.) It's also a very oddly contoured concern: for instance, we only count *some* things as 'authored' (papers, but not signed checks or letters of recommendation, will turn up in someone's *Gesammelte Werke*). Foucault is suggesting that the contours of and the concern with authorship more generally are the products of a random historical walk; once we understand the rough shape of this history, we are going to find ourselves wondering, Why care about this? Turning back to attitudes and actions, the Foucauldian thought would be: why should we build an equally contingent analog of this historically recent fixation into our theory of personhood? Why should we *care* if our attitudes or actions are 'owned' or 'authored'?

idea, satisfaction comes out as a consistency requirement on the distinctive sort of policies that meet the other requirements on Bratman's list.[23]

However, it's unrealistic to expect a great deal in the way of otherwise-Bratmanian policies meeting the satisfaction condition. ("Otherwise-Bratmanian": policies that meet all the conditions on the list except that one.) Suppose that, as a child, you copy your policies from your environment. In that case—but pretty much any plausible childhood alternative to mere copying will share this feature—you shouldn't expect consistency. Why should your parents, the TV set, and your childhood peers all have the same, or even compatible policies? If you are to render the randomly copied policies consistent by sorting them out, *you* sort them out, in which case the account helps itself to the very type of attribution it's trying to explain, before the conditions for such an attribution could have been put in place, and so is viciously circular. But if you can't sort the inconsistent policies out, fully Bratmanian policies will be rare: so rare, that if they are what account for superlative ownership of attitudes, valuing, and autonomy, these latter will themselves be too exotic to call for much in the way of philosophical attention. (Call this second problem the *genetic objection* to satisfaction.)

Third, the desirability of satisfaction should not be taken for granted. Remember the nineteenth century's *Sturm und Drang* movement; Romantics worried that if you didn't have different attitudes struggling to correct one another (that is, if you weren't dissatisfied with yourself, in both Frankfurt's and Bratman's variants of the notion), your personality would be static, and you would be incapable of maturation and growth. (Call this the *Romantic*

[23]The notion is taken over from Frankfurt, 1999, ch. 8, where it means that one is not disposed to alter one's psychic state. For Bratman's adaptation, note the "first" and "second approximations" on p. 35, and the "final modification" on p. 44. Why do it this way? Because it's easy to require that a superficial bad attitude won't do (say, a mere desire not to have the policy, one that amounts to no more than a whim). Once we require that the bad attitude be at a level of psychological depth similar to the one whose superlative attribution it undercuts, then it has to be another Bratmanian policy functioning to bind together the various temporal components of an agent—or a quasi-policy playing a closely analogous role.

When defining 'satisfaction', there is an adjustment we will want to make to Bratman's rendering. Most of the time, having one policy amounts to having a bad attitude about another because they conflict in the obvious and direct sense: the policies tell me to do different things. But policies can also *undercut* one another, without conflicting, as when one policy tells me to pay no mind to gossip, and also to do such and such, while a second policy, which I picked up on the gossip circuit, agrees with the first that I should do such and such. In my view, Bratman-style satisfaction is best construed as requiring not just that one have no further policies that conflict with the policy in question, but that one have no policies that undercut the policy in question.

objection to satisfaction.)[24]

I do not want to pursue these objections as they are. Our observations—that we are unlikely to have many deliberation-governing policies with which we are satisfied, that there is a history of thoughtful nervousness about the settled and static personality, and that we do not have an argument for Bratman's Master Move—will turn out to be dry runs for more pressing complaints anchored in an account of humanity's shared adaptive strategy. A correctly functioning personality will not only often be unsatisfied with its otherwise-Bratmanian policies. It will be so especially when we understand the agent to be thinking *for himself*.

3

Psychological states and psychological structure characteristic of an organism are ordinarily to be understood as part of the organism's strategy for responding to challenges it faces. To be sure, this claim has to be qualified in many ways: natural selection does not always produce cleanly engineered, optimal solutions to adaptive problems; in culturally plastic creatures (i.e., us), not all psychological structure is a product of natural selection. Nonetheless, an organism's psychology should make sense as part of the organism's life. A trivial for-instance: plans and policies control behavior over extended periods of time, and so a creature that lived only for moments—while it *might* have a mental life—would have no occasion for either plans or for policies.[25] In assessing Bratman's proposal, the first

[24]Frankfurt is clearly aware of the objection, and he goes to the trouble of denying the Romantics' claims (1999, p. 102)—but not of providing a convincing argument against them. (Margaret Bowman helpfully pressed this concern.)

[25]Bratman's work can be understood as a sophisticated descendent of Nagel, 1978, one which takes up its discussion of the metaphysical interpretation of the self, and so clarifies the deep connection between temporal extendedness and practical rationality. Where Nagel tells you merely that what it is to be a temporally extended creature is for future reasons to give you reasons now, Bratman provides a great deal more in the way of such a metaphysical interpretation: *these* temporally extended creatures have reasons with weights set by policies that... etc.

In recent and related work, Ferrero, 2010, has offered Bratman an account of the stability of intentions: your intentions are stable because you've delegated the decision to your past self, and you're pretty sure he made the decision the way you would've. However, Nagel persuasively argued that internalist or present-aim-theory approaches to prudence are misguided: it's a mistake to look for a reason for me-in-the-moment to care about my reasons at other times; if I-in-the-moment need one, I'm not really a temporally extended agent at all, but rather merely a community of time slices. Ferrero's view is internalist in spirit, in that it attempts to give you a reason, one that makes sense to your momentary

question to ask is: What kinds of creatures are *we*?

We survive, and often thrive, in a world that is, albeit at intervals, deeply surprising, and by 'deeply,' I mean that our collective experience teaches us that there's pretty much nothing we can't be surprised about. We get along in our surprise-laden environment by availing ourselves of a distinctive capacity for ecological specialization. Where other species become specialists in hardware, so to speak, reshaping themselves over evolutionary time to have, say, the long neck needed to reach those upper branches, we do our specialization in software, and the software allows much more extreme specialization than one tends to see otherwise: call it *hyperspecialization*. What is more, the software is reprogrammable, which means that, over and above the collective human capability of occupying multiple niches in an ecology, an individual human can occupy them one after another.

Now, parts of our environment are, for sometimes quite long stretches of time—in suitable respects, and after a while—*un*surprising. In making them serve us as niches, we identify their relatively stable features, and develop representational schemas, systems of standards and methods of deliberation that allow us to occupy and to exploit them. But, and this happens for many reasons, humans often have to move on: to exit one niche, and either go looking for another, or learn to manage in parts of the world that do not have the stability or the structure to amount to a niche.

Over time, we have accumulated a large inventory of attempts to complete the sentence, "Man is…" (You know, "Man is the animal that laughs"—slogans like that.) My label for this strategy is being volunteered as another contribution to it: *Human beings are serial hyperspecializers.*[26] Our present question, evidently, is what sort of agents serial hyperspecializers are going to turn out to be.

This intentionally biologized description of human life requires an example if it is not going to be misleading; because humans occupy many of the niches in the ecologies that they—after enough time—more or less constitute, for humans, ecological niches are often indistinguishable from social roles. Since the audience for this paper is likely to consist for the most part of philosophers, here is the close-to-home illustration: philosophy, seen as a profession or social role. (Here I want to leave open the question of whether that is a good way to think of philosophy; suffice it for the present that

self, to heed your past self's decisions. Evidently, the fit between Bratman's project and Ferrero's analysis is awkward; Ferrero's analysis amounts to giving you further reasons to act on your previous decisions, where the propensity to *just* so act, without such further reasons, is a structural feature of creatures whose lives are extended through time.

[26]For further discussion, see Millgram, 2009b, Millgram, 2009c, and Millgram, 2009a.

the professionalization of the activity is widespread and familiar, whatever its merits and demerits may be.) Philosophers today take it for granted that their potential employers are, almost all of them, institutions of higher education; that publication is required; that venues for publication include professional journals, books bearing the imprint of an academic press, edited collections of essays, book reviews and—very recently—blogs. (These are examples of relatively stable features of the environment that define a niche.) In the US, there is a widely accepted institutional pecking order, defined by someone's top-fifty list; there is a slow-to-change map of areas of specialization, that is, subfields such as philosophy of mind, history of modern philosophy, or philosophy of biology. (The first is an example of a system of standards adopted to make the niche navigable; the second is an example of a representational schema used to navigate the niche.) Against this sort of background, the deployment of plans and policies makes sense. Plain, ordinary plans might include the intention to specialize in such and such an area, to publish in such and such journals, and to write an academic book. An example of a Bratmanian policy might be the intention to count Leiter Report rankings as overriding considerations wherever they are relevant.

To reiterate, human life is an uneven mixture of the routine and the novel, and individuals may need to leave a niche—sometimes abruptly, and for reasons they could not have anticipated. In Germany, in the 1930s, philosophy professors did not read the Leiter Report, so let's imagine one of them whose Bratmanian policies include accepting the advice of his Doktorvater, not making life choices on the basis of outlandish rumors, and not letting politics impinge on career decisions. Suddenly, this philosopher discovers that he can no longer hold an academic job, that he can't use the general post office or employ his housecleaner, that his *name* has been changed, and that he has to wear a gold star. His Doktorvater tells him not to take it so seriously, that times are tough but that they're bound to get better, and that while the ruling party is overdoing it in his case, nonetheless, one has to admit that something had to be done about Jewish control of German cultural and economic life. The philosopher is hearing outlandish rumors about what is happening to his arrested and 'resettled' coreligionists. Overall, politics has come to pervade his life in ways he would have found unbelievable only ten years before.

As the illustration reminds us, working up plans and policies for genuinely unexpected change is at best an exercise in futility, and at worst incoherent: if you don't see it coming, you can't have a plan for it. (Even if you are in the unusual position of being able to enumerate all the possibilities, when there are too *many* distant possibilities of this kind, you can-

not formulate well-considered contingency plans for all of them.)[27] Serial hyperspecializers—us!—sometimes (not always: there are obviously other reasons as well) abandon the niches they are in because they are faced with genuinely unexpected change. So they often are not acting on plans or policies when they exit niches, and bear in mind that in dramatically new circumstances, it takes a while to come up with a reasonably adopted plan or policy; in such circumstances, we should expect that, for a while at least, their actions are not policy governed. Psychological structure, we observed, should fit into the shape of the life of an organism. So we should expect to find the psyches of serial hyperspecializers to be a mixture of plans or policies (suitable for use in relatively stable niches), on the one hand, and psychic equipment for coping with the impossible-to-anticipate, on the other. An account of agency that is made out solely in terms of plans and policies can be no more than half of the theory we are after.

If the structures of (and strictures for) agency match the shape of an organism's life, if humans are serial hyperspecializers, and if the life of a human being is normally segmented—a philosopher for so many years, then a party journalist for a period, then a media consultant, etc.—we should expect to find that human beings are, always potentially, and frequently in fact, *segmented agents*.

4

At this point, the skeptical reader may be wondering whether our critical assessment of mainstream accounts of agency presupposes too much in the way of a controversial and very-big-picture recharacterization of humanity. So let me emphasize that the problems to which the recharacterization directs our attention are problems by the lights of ordinary common sense. The perhaps overly dramatic (but nonetheless real-life!) illustration was meant to make vivid the point that we can be faced with the genuinely impossible-to-anticipate, and therefore with circumstances for which we cannot have

[27] The point that you can't prepare for what you can't anticipate applies not only to policies but to quasi-policies, and here is a place to mark a complaint readers of Bratman often come up with, but which we can put aside. McDowell, 1998, pp. 57f, has observed that not all of our intentions are 'codifiable', whereas Bratman's policies seem to amount to codification. I want to allow that Bratman-style policies need not be fully articulated, or even articulable by an agent who has them. But we should not mistake action-guiding attitudes which you are unable to spell out for attitudes capable of providing thoughtful guidance in situations the like of which you have never encountered, and in which you have never invested any thought. That you cannot say what your policy is does not allow your policy to violate the law of conservation of cognitive work.

prepared suitable plans or policies. A fortiori, we cannot have prepared suitable Bratmanian policies. Sticking with the example, it is clear in retrospect that the right deliberative move for the 1930s German-Jewish professor is to drop his Bratmanian policies like a hot potato: those who didn't (of course, along with most of those who did) ended up being packed into cattle cars and shipped off to slaughter. Now, Bratman insists that he "does not mean [Bratmanian] policies [to be] immune to rational revision. . . [his] project is not to describe some irrevocable foundation at the bottom of all further practical reasoning."[28] But suppose the professor *does* drop his Bratmanian policies: can this be, by Bratman's lights, autonomous action, something that *he* really *did*?

Bratmanian policies are reflexive: they apply to themselves. So one can "reflectively reassess and revise where one stands" with respect to one's Bratmanian policies—*if* one has a Bratmanian policy, triggered by the conditions in which one finds oneself, that tells one to revise one's policies in those conditions.[29] Although there are delicate issues to navigate, about how to individuate policies, and when one can square revising one policy on the basis of another with the satisfaction requirement on Bratmanian policies, we ought to allow that policies can be surrendered on the basis of an agent's further policies. (The ship which Neurath described as being rebuilt at sea, plank by plank, is familiar philosophical shorthand for this point.) Perhaps I cannot drop all my Bratmanian policies at once, but can drop any one of them, as long as there are at that time other and suitable Bratmanian policies that I am retaining. And eventually, revising my Bratmanian policies one by one, I can revise them all. What is more, policies can be rough-hewn and broadly framed, with an eye to problematic circumstances: "In an emergency, stop, look and listen." All that granted, if a policy is not so broadly framed as to be entirely content-free, there will nevertheless be

[28] P. 36.

[29] *Ibid.*; in his early work, Bratman introduced plans as stable in this sense: it takes special circumstances to make one reconsider them. (See pp. 289f for an attempt to derive a commitment to taking the means to your ends from the requirement of means-end consistency together with stability: stability is thus being treated as a very deep feature of practical rationality.) At the time, he treated the disposition to reconsider a plan as not itself further analyzed, as determining, roughly, *how much* it would take to make you reconsider, and as subject to roughly consequentialist assessment: the disposition, along with the setting on its dial, and thus failures to reconsider plans on particular occasions, could be justified by showing that on average the results were good. Bratmanian policies amount to an alternative to both the 'volume setting' model—they allow finer-grained triggers for reconsideration—and to the consequentialist mode of assessment; they are thus an advance on the earlier position.

circumstances it does not cover. And that is not a bad thing: because one cannot have given thought to unanticipated circumstances, a policy that did cover all circumstances would be a thoughtless—a *foolish*—policy to have. Running with that last proposed policy, not everything that requires rethinking what you are doing is an emergency; or alternatively, you cannot be equipped to recognize everything that counts as *that* sort of emergency; in any case, whatever we are told in grade school, stopping, looking and listening is an appropriate response only to *some* emergencies.

I have found that philosophers, exhibiting what seems to be a professional reflex, respond to illustrations of unanticipated circumstances by trying to think of policies that would handle them. The response is common enough to justify taking time out to explain why it is misguided.

First of all, the order of the quantifiers matters. Even if it were true that, for every circumstance you might face, there is a policy you could have that would cover it, it would not follow that there is a policy that would cover every circumstance you might face. I have already suggested that, because you can only have put so much deliberation into any policy, adopting a policy that covers all bases is almost inevitably a bad idea.[30]

Second, such an all-purpose policy is supposed to manage an intelligent response to whatever circumstances you encounter. Policies amount to collections of rules launched by specific trigger conditions (if you have a programming background, a policy is a lot like a Lisp `cond`). And that allows us to recognize the policy these objections are after as the holy grail of 1980s AI, the program of machine intelligence through rule-based systems. Some very bright people were unable to make this approach work; if the MIT Artificial Intelligence Laboratory failed to find the policy you are imagining, why be so confident it's there for the having? There was a reason that the field abandoned the approach and moved on.

Third, and last for now, even if a policy that covers all the bases exists in

[30]See Millgram, 1997, ch. 4, for an argument that an agent will not be able to live by a contentful plan that is chosen without taking into account experience of the sorts of circumstances the plan covers.

And here is a quick argument that you cannot in fact have a policy that covers all bases. For serial hyperspecializers, a policy that covered all bases would have to cover activities within specialized niches. It is characteristic of these specialized niches that their occupants develop descriptive vocabularies needed to navigate the niche, and that these are not intelligible to outsiders. The trigger conditions of policies that govern intelligent behavior within such a niche will have to be largely formulated in the specialized vocabulary. Since anyone is an outsider with respect to most specialized niches, no one can so much as understand a policy that covers all bases. You cannot have a policy you do not so much as understand. Consequently, no one can have a policy that covers all bases.

the Platonic Heaven of the Forms, it's clear enough that most agents don't know about this policy and haven't adopted it. (If they had, surprise would itself be a surprising thing.) A theory of agency should not restrict itself to agents who have adopted some policy that already solves all the challenges of agency, in something like the way that a methodological proposal in philosophy of science should not restrict itself to the scientists who have already arrived at the final Theory of Everything.[31]

Returning to the illustration at hand, I constructed the example so that the German-Jewish professor's circumstances do not trigger policy-driven reassessment of his Bratmanian policies, and that is a legitimate stipulation because one's circumstances can be impossible to anticipate, and so one's Bratmanian policies may well not be tuned to the conditions in which, at one juncture or another, one finds oneself. If we are considering such an occasion now, and if the professor in the illustration does drop his Bratmanian policies, it is neither by virtue of their reflexive application nor by virtue of their Neurathian application one to another, and therefore, it is not by virtue of applying his Bratmanian policies at all.

Recall that Bratman, unlike most of the competition, does not insist that his account of attribution and autonomy lays out the only way an action can be self-governed and fully yours. So what we can say is that Bratman's account does not show us *how* the decision to drop a Bratmanian policy in the face of genuinely unanticipated circumstances can be, in the full-fledged sense, your decision, and how actions pursuant to that decision, but taken before replacement Bratmanian policies are formulated, can be autonomous action.[32] Even without bringing to bear the characterization of human beings as segmented agents, we can see that something has gone badly wrong: being sent like cattle to the slaughter is the very opposite of self-governing action, and if and when the German-Jewish professor ditches his Bratmanian policies and flees the country, he is, as we say, taking matters into his own hands and thinking for himself.[33]

[31] Christine Chwaszcza has suggested to me that perhaps the bias towards full-information accounts of preference and the like is best explained as an inheritance from an older philosophical tradition that attempted to adopt the point of view of an all-knowing God. If that is right, once we have left behind the theological interest in God's plans for us, we should also drop the methodological orientation towards (what we can call) hyperintentionality derived from that interest.

[32] Thus when the argument, suitably reformulated, is directed to the competition, the upshot will be that when the German-Jewish professor swerves away from his former 'constitution,' higher-order desires, and the like, that choice *cannot* be autonomous, or fully attributable to him.

[33] Chesnow, 2008, p. 16, describes his parents' postmortem of their insufficiently proac-

5

Return to the ecological characterization I have adopted of our species. Segmented agents are a psychological adaptation to a life consisting of stints (both longer and shorter) in ecological niches that typically are also social roles. These niches are characterized by relative stability, because without stability, it would be impossible to have that part of the ecosystem serve as a niche. But there is no reason to suppose that there is similar stability everywhere else, and plans and policies only make sense against the background of a relatively stable environment. So we have to adopt a more nuanced view of the role of plans and policies—and thus, of Bratmanian policies—in human life. Intelligently formulated Bratmanian policies can be suitable guides for action *within a stable niche*. (With, however, an exception: an agent currently occupying a niche may need to think outside the box when deciding whether to push the eject button.) Such policies are not, evidently, suitable guides for the big, wide world, but, I want to insist, in *Structures of Agency* Bratman has done us an important service, that of diagramming psychological structures, and distinctions that accompany them, that serve us well in well-structured and well-understood environments.

Sometimes we're invited to think of ourselves as space-time worms, but worms don't have spines, psychic or otherwise; snakes have spines, so maybe the Bratmanian picture is an invitation to think of ourselves as space-time snakes. But we are not space-time snakes. Segmented agents have (roughly, more or less) one Bratmanian spine per segment: psychic structure that manages one's agency during the time one occupies a particular local environment stable enough to serve as a background for plans and policies. Let me emphasize that that's a first approximation: people can have a foot in more than one niche at a time, and they can spend extended periods of time outside of the constrained environments that serve as niches (in which

---

tive response to their similarly changed circumstances. His father ran a business in one of the Baltic states that was doing very well on military contracts in the runup to the Second World War, and failed to take advantage of opportunities to emigrate. Chesnow recalls his father recounting Kipling's instructions for trapping a monkey: you put a banana in a cage whose bars are wide enough to let the monkey's spread hand in, but not wide enough to let a fist clenched around a banana out. "My father," Chesnow continues, "finished this story with the words, 'I'm the monkey.'" When you can't give up your policy, except for reasons that the policy itself antecedently specifies, it's all too easy to end up being the monkey: Chesnow's parents did not survive the war.

As the example reminds us, we do not always respond to changed circumstances in the way they seem to demand. A lower-key example (due to Elias Moser): we've learned the hard way that relocating at-risk youth into a different environment, in the hopes of prompting new Bratmanian policies, is often unsuccessful.

case, they take on the ecological appearance of weedy species like rats or cockroaches, and may fail to have Bratmanian spines at all). To say that each segment will normally have its *own* psychic spine is to say that these backbone segments are not normally connected one to the other in the way that their components are connected to one another: that is, one segment's spine is not joined to another's by plans or by policies. As a placeholder, we can say that they are held together by being all embedded in one temporally extended person.[34]

With this image in mind, we can say a bit more about why we should not be treating a model of a time-bounded segment of agency as a model of human agency over the course of a life. One challenge that human beings face is that of managing the transition from one segment to the next. Coping with this challenge can't simply be a matter of deploying psychological machinery that governs the segments; that machinery is effective because it can exploit the stability and constrained environment of a temporarily occupied niche; but stability and constraints on which one can rely are just what is missing in the transition from niche to niche, and in the world at large. So reliance on Bratmanian spines must alternate with turns to a very different form of

[34]Vogler, 2002, pp. 106f, complains about Bratman's earlier work that planning theory is suitable for the managerial classes (and only the contemporary version of them, because, centuries ago, managers didn't own dayrunners or scheduling applications). But it is not suitable for anyone else, and so it cannot be a satisfactory account of practical rationality, which has to be rationality for everybody. We are now in a position to amend Vogler's complaint: plans and policies are indeed usable, and not just by the managerial classes, but in most niches: plans require, in the first place, stability, and if it's not stable, it's not a niche. There is something her objection is getting right, however: plans and policies work well neither in the large part of human life that is lived outside one or another stable niche, nor during transitions from niche to niche. That doesn't show that planning theory isn't part of the theory of rationality; it does show that it can only be part of the theory, because we need to understand what rationality amounts to when you are thinking outside the box.

Does my description of the human *ergon* imply that a well-lived life is spent dealing with emergencies? As I rendered the human species form, it might seem that the happy life must be spent leaping from one specialized niche to another, and that if you're not constantly throwing everything you cared about overboard, you're a failure. But surely a life without such upsets is the *better* life, and the one to which we should aspire.

That your *ergon*—your design description—involves specifically *serial* hyperspecialization doesn't imply that if you don't use the capability, your life is thereby unhappy. The design solution that human beings implement includes many features that might never be used in a well-lived life: just for instance, the ability to metabolize your own muscle tissue when you're starving (or also, when you're doing exercise you haven't trained for properly), or the ability to learn more than one language natively. (Thanks to Michael Millgram for help with the first example; for second thoughts about applying the concept of *eudaemonia* to serial hyperspecializers, see Millgram, 2009c, sec. 6.)

practical rationality.[35]

The problem I am posing has an analog in the philosophy of science. It is now a familiar doctrine that, during periods of normal science, scientists proceed on the basis of policies that determine what counts as successful argumentation, legitimate results, and so on; these policies express what counts as scientific rationality during that period. But during so-called scientific revolutions, those policies have to be abandoned and rethought from scratch.[36] Because the rethinking cannot be conducted in accord with the policies that are being jettisoned, it has been hard for both philosophers and historians to see how the outcomes of scientific revolutions can count as rational. Consequently, it has also been hard to see why scientists should get credit for them: should an Einstein get a Nobel Prize if his achievement was really just a bit of lucky craziness? But when Kuhnians treat paradigm shifts as irrational episodes sandwiched between the longer stretches of scientific rationality, rather than as among the highest intellectual achievements of the scientific tradition, achievements that stand out as the strongest examples of rational thought that we have available, that is evidently a theoretical failure on the Kuhnians' part. It is a philosophy of science suitable for a plodding and intellectually crippled version of the enterprise that might be conducted by a more limited species, one incapable of invention that was both genuine and thoughtful.[37]

[35]There is an important class of exceptions that I want to register: sometimes the niches are sufficiently entrenched, and the transitions between them sufficiently standardized, for there to be metapolicies governing transitions between niches and the Bratmanian policies appropriate to them. For instance, an employee may be governed by a Bratmanian policy at work which we can abbreviate as: be professional. At home, he is governed by a very different, much more domestic Bratmanian policy. His metapolicy is to switch from one to the other during his commute. Notice that this metapolicy is underwritten by the stability and standardization of the arrangement: indeed, at this point, the segregation of the policies is supported by very-hard-to-change physical infrastructure, the differently located residential and industrial neighborhoods between which the employee must drive.

Obviously I think it's a mistake to understand the segmentation of agency as a sort of super-policy. But notice that even if it could be construed that way, for most of us it would not count as a Bratmanian policy; most people are unaware that their lives are segmentable, and in their case, the impossible super-policy would violate Bratman's transparency condition.

[36]Kuhn, 1970.

[37]Friedman, 2001, suggests that thinking outside the box is *philosophy*, and that's why awareness of philosophical developments has an important role in scientific revolutions. While there is certainly something to the suggestion, not all thinking outside the box is philosophy; when a lurch in your career puts you in the position of having to figure out how you're going to reinvent yourself, your deliberations may well not count as philosophy by anybody's lights.

It takes a big man to admit he's wrong, or that's what popular wisdom says. The problem we were homing in on is that, on a view like Bratman's, you can't ever be that big a man, because you can't admit you *really are* wrong, wrong all the way down: not wrong in ways for which your policies leave room, not wrong in ways that are made out through the reflexive application of self-governing, reason-determining policies, but completely wrong, wrong even about what would be a good reason to change your mind. Or rather, the problem is that coming to that point can't be something *you do*—though it can *happen* that you come to see your former attitudes as wrong, all the way down.[38]

Once the point is conceded for the dramatic illustrations, it should be allowed in more mundane cases of niche-switching as well. When we show, in a laboratory experiment, that certain birds are sensitive to the Earth's magnetic field, we correctly assume that that sensitivity is deployed outside the laboratory; if we convince ourselves, by considering artificial or extreme circumstances, that human beings are able to manage choice that is not policy driven, we should similarly assume that they will deploy that ability wherever there is an advantage in doing so.[39] A philosopher may decide that the business of philosophy is just not for him (and we philosophers can all think of people we know who did decide that, and decided on the basis of reasons for which their reason-governing policies had left no room). When he does, he will slough off the Bratmanian policies that guided him as a philosopher—perhaps that system of standards taken over from the Leiter Report—and doing so, without a Bratmanian policy to guide his deliberations about the matter, can be the most difficult, demanding and

[38]Here Bratman is, again, representing mainstream commitments that are largely implicit, but occasionally explicitly acknowledged. For instance, Watson remarks: "The important feature of one's evaluational system is that one cannot coherently dissociate oneself from it *in its entirety*... One can dissociate oneself from one set of ends and principles only from the standpoint of another such set that one does not disclaim" (2004, p. 26).

[39]And indeed, as Jenann Ismael has reminded me, the phenomenon I'm identifying is built into the completely routine formal structures of our ordinary early lives. First you go to elementary school; then to high school; then to college; then perhaps to professional or graduate school; finally you emerge from this series of cocoons and get a real job. At each stage, your policies, goals, preferences and self image have to be reworked pretty much from scratch. It's not just that an elementary school pupil who tried to formulate a way of getting through the world intended to guide him through the subsequent stages would be very peculiar; even if he tried, he would be badly served later on by sticking with the plans and policies he'd made up as a child. Later life is in fact no different from childhood; as you mature, you have to rethink your Bratmanian policies and other such components of your personality and—we can hope—you're never done maturing.

admirable form of self-government at work.[40]

Bratman allows that there may be psychic structures which are not policies, but which play much the same role that his Master Move assigns to policies; he calls them quasi-policies, and mentions ideals as an example. So let's just check that the point we're making sticks even when we have Bratman's quasi-policies in the mix. One very striking aspect of the human predicament is that ideals often have a shorter shelf-life than people: call this the *Shelf Life Problem*. (Think of the ideological politics of the twentieth century, during which communists outlived the relevance of communism, and Germany's 'revolutionary nationalists' outlived the respectability of their movement.[41]) It is unusual that such an ideal contains within itself the intellectual and emotional resources for recognizing that its time has passed; on the contrary, well-developed ideologies generally provide catch-all methods of dismissing the reasons adduced against them.[42] When someone

[40]Let me add a qualification to my endorsement of Bratman's account as a satisfactory reconstruction of how we manage within a stable niche. Because human beings are opportunistic when it comes to exploiting available resources, cognitive and otherwise, we should expect that when we focus tightly enough even on activities within stable niches, we will also find action not fully controlled by Bratmanian policies. Let's revisit the very low-key example used by Bratman, long ago, to introduce plans, and which we mentioned in note 17. Once again, you are about to drive up from Palo Alto to San Francisco, and you could take either of two freeways, 101 or 280; but now, let's imagine that you have adopted a Bratmanian policy covering such decisions, namely, you will choose the route based on the time you expect it to take, with stop-and-go traffic as a tie-breaker. Let's further imagine that, in this case, 101 and 280 are on a par, as far as your policy is concerned; if you simply decide to take one or the other, it will have been full-fledgedly your own choice. But now, suppose that, in these circumstances, you form an intention to take 280 because it strikes you that the scenery along 280 is much nicer. This is introducing a further consideration, one not covered by your Bratmanian policy. So is the ensuing drive up superlatively *your* action? My own sense is that, in circumstances like these, the conceptual apparatus should be thought of as delivering a mixed answer: it is your decision, up to a point. And such mixed outcomes need not arise only in trivial circumstances: a physician, for instance, is only too likely to have to make choices that, while they conform to his niche-specific policies, also outrun them, and in just this way. (I'm grateful to Paulina Sliwa for this last example.)

[41]The label comes from Rose, 1990.

[42]Unusual, but nevertheless, some ideals exhibit the phenomenon Nietzsche calls "self-overcoming," in which the rigorously applied ideal is turned against itself. What matters for the present point is, first, that self-overcoming is not (as Nietzsche seems to have thought) the fate of every ideal, but only of some of them. And second, the criticisms an ideal is able to direct against itself are often not the ones that matter: the ideal of fairness may motivate a policy that you are later able to understand to be unfair; but perhaps the readjustment required in the face of devastatingly unfair damage to your prospects is accepting that sometimes you have to let go of the preoccupation with what is fair, and get on with what is left of your life.

(as many, perhaps most of us, will have to) lets go of an ideal that has reached its expiration date, that is, when it is done right, their choice, in the fullest-fledged sense, and autonomy at its best; but it is not, usually, a choice that is guided by the ideal itself—or, for that matter, by a distinct and previously formulated policy for dealing with one's ideals.

Recapping, and spelling out the problem a little more slowly, allow that it is a frequent enough requirement in the lives of segmented agents that they have to admit they were wrong. Allow also that this sort of reconsideration can be understood as autonomous, and the attitudes involved in it, as superlatively your own. If the only way we now have on board to understand superlative attribution and autonomy is made out in terms of Bratmanian policies, then to accept the Bratmanian account as adequate is to presume, for practical purposes, that such reconsideration can always be managed by one's Bratmanian policies. (If you didn't accept that, you'd make room in your account for full-fledged choice of other sorts.) That in turn is to presume that we never have to reconsider Bratmanian policies in ways those policies themselves don't license, which is to presume that there are certain things we can't be wrong about: as I was phrasing it earlier, that we can't be wrong all the way down.

Once again, the view we are considering is being taken up as a best representative of a large and mainstream family of views. And you might be wondering whyever a philosopher would get himself into the position of assuming, even implicitly, that you can't be wrong all the way down: the supposition is strongly belied by ordinary experience.[43] Rather than take up this diagnostic question, let me turn briefly to the task of replacing obsessive preoccupations with useful devices, and to proposing a real function for the distinction between attitudes that are *merely* attributable to you, and attitudes (and choices) that are *superlatively* attributable to you.

[43]While it's easy to wonder whether one or another version of noncognitivism is in the background, Bratman intends his treatment to be metaethically neutral (pp. 174, n. 29, 226, 297, n. 26).

I am insisting that you can be wrong, completely wrong, wrong all the way down—and that, for human beings, failure to acknowledge this has the distinctive look and feel of self-deception. And isn't this just skepticism, towards which we should also be considering a diagnostic posture?

Not really. The iconic skeptical suggestion is that you *might* be wrong about everything (not that you really are), and if you were, you would never know. Philosophers' skepticism, the modern version anyway, is an intellectual enterprise where the live-but-not-actual possibilities are used to reassess the epistemic status of your beliefs. Whereas I am reminding you of something that grownups already know: that you *are* wrong about a very great deal, and that you're all-too-likely to find out, probably in ways that amount to hardship; the epistemic status of your beliefs and other attitudes is beside the point.

6

A segmented agent can come to realize that its policies (and a fortiori, its Bratmanian policies) are thoroughly wrong, wrong all the way down. That means that it must be cognitively equipped to tell when its policies are bumping up against a recalcitrant reality, but we can be pretty sure that a segmented agent's equipment for telling that things have gone wrong in a way that requires massive revision *won't* be a policy. Things go wrong in unanticipated ways, and there are no criteria for telling when *that's* happened, and no rules for responding to unanticipated circumstances. Policies just set criteria or trigger conditions for rules, and consequently, you can't have a policy for handling the unanticipated.

In those circumstances, a person will have to rely on *himself*, rather than his policies: from which it follows that the distinction between himself and his policies must make sense, and make sense to him. Returning for a moment to Bratman's Master Move, this is a reason to resist the assumption that when your policies speak, *you're* speaking. The need for that distinction should not be taken for an occasion to revert to the untenable picture of an immaterial self that is different from and concealed behind the individual's psychology. I will presently take up the question of what the third way here has to be.

I earlier recommended reopening the question of why we care (and whether we should care) about superlative attribution—and, by implication, the distinction between autonomous and nonautonomous action. However, before making my own suggestion on this score, I need to register a caveat. These are starting to seem like useful concepts (as construed by Bratman, but, yet again, I take him to be giving the best available rendering of a widely shared view), when applied *within* the local framework of a niche. *Outside* those niches, it's less obvious that these are the most interesting or important contrasts and distinctions to pursue. This is perhaps especially striking in over-dramatic examples of the sort we were considering. On the one hand, even a self-aware agent is likely to be overwhelmed by such circumstances; autonomy is a concept used to assess the quality of one's choices, and when one is swept away, one often has, as we say, no choice. On the other hand, when people rise to the occasion, even in such circumstances, we say that, then especially, they are thinking for themselves. Evidently, the problem with the sorts of account we are considering is not just that they don't provide the wherewithal to explain how the German-Jewish professor's decision to ditch his prior Bratmanian policies can be *his* choice, but that they require that attitudes, choices, and actions fall on one side or the other of

distinctions that they rather seem to straddle when agents are performing well in unstructured and challenging conditions.

So the point of the argument we have been constructing is not that we have the view that certain responses to real or imagined circumstances are autonomous; that Bratman's account does not classify them as autonomous; and that that is an objection to the account. It is rather that we *need* intellectual equipment with which we can assess performance in a hiatus between agential segments, and that the notion of autonomy as construed by Bratman (and the other philosophers for whom I am using him as a representative) does not meet that need.

Now, suppose a segmented agent has realized that its former policies, Bratmanian and otherwise, have to be replaced. Since we are using Bratman's account as our foil, we can assume that its psychology mostly consists of a mass of ingrained and habitual policies, attitudes that are endorsed and underwritten by such policies, and other psychological structure that has accreted around the relatively stable Bratmanian policies that managed activity during a previous segment of its life. All of that won't go away in a moment, and at the onset of the changeover, a newly-invented replacement policy will be no more than a tiny, not-very-well-entrenched wiggle in the psychic stew. Successful changeovers will require self-monitoring, to allow a segmented agent to catch itself when it goes on doing things the old, habitual way. So a segmented agent needs to be able to say, in what we can think of as a proleptic or anticipatory register: No, that's not *my* policy anymore. This is *a* function that the distinction between merely mine and superlatively mine serves; I strongly suspect that it is not its only function (and so, this is only a first step in addressing the question of what our practices of superlative attribution are really for), but it is at any rate a job that has to get done.[44]

Notice that the proleptic form of the superlatively-mine/merely-mine distinction is deployed only when one is, in the sense we introduced earlier, *dissatisfied*: whereas recall that Bratman imposed satisfaction as a precondition for making the distinction at all. Here's what's at stake (returning to and upgrading the Romantic objection): for segmented agents, dissatisfaction is a *normal* component of a *successful* life. Consequently, and leaving aside the technical sense of "satisfaction" for a moment, it is not something

[44]A good way to see that it's not the only use to which we put such distinctions is that even when one correctly says to oneself, No, that's not what I think! one may later not get to say, to someone else, of the very same attitude embedded in the very same episode, No, that's not what I thought—or anyway, not with the same definitiveness. (I'm grateful to Tom Pink for the observation.)

about which they should be dissatisfied; when you rise to the occasion, and the occasion requires you to switch gears in unanticipated ways, you may well feel, not dissatisfaction, but appropriate pride in taking a stand against older and perhaps deeply entrenched attitudes. To switch gears in this way is to come to contain competing and typically independent structures of agency. These structures—whether construed as Bratmanian policies, or as quasi-policies, or as 'constitutions', or as hierarchies of higher-order desires—are *normally* conflicting and mutually undercutting. Therefore, during the period in which they overlap, the agent is (reverting now to the technical sense) dissatisfied. But that's (often) a good thing, and the agent can be entirely aware that it is.

7

We have been criticizing Bratman's view as a best representative of a philosophical approach that consists in looking for psychic structures with this feature: when they guide, you govern.[45] If I am right, we need to be taking a very different approach: that of making philosophical sense of forms of receptivity to the world such that, when the *world* guides, via those forms of receptivity, you govern. I do not have the account I am pointing towards in my pocket. But I do want briefly to address two sorts of incredulity that are likely to meet the proposal. How can it be *self*-government when the guidance is coming from outside of you? And how can the *world* be providing guidance? Why isn't that suggestion a return to superstition, and the pre-Enlightenment magical thinking that takes decision-making to consist in the search for omens?

You can't generally understand or assess creatures outside of the environment in which they function, and if that is true of humans, then assessments of agency that consist just in looking *inside* a person, at his internal psychic structures, will get off on the wrong foot. If that is true generally, we should expect it to be true of assessments of autonomy, and of the superlative attribution of attitudes and actions. Philosophers with the family of views that I am resisting sometimes analogize the person and the state, so perhaps it will help to remind ourselves that my recommendation *is* our practice with respect to states: we judge that Neville Chamberlain's England "slept," not because its internal structures of governance had changed, but rather because the British political system failed to respond intelligently to an unprecedented external challenge. That near-fatal lapse in Britain's collective

[45] This way of phrasing the objective comes from Bratman, 2009, p. 430.

autonomy was a matter of failure of receptivity: an inability to acknowledge and act on observations for which the prevailing policies left no room.

Returning now to individuals, examining organisms together with their environments allows one to pick out signaling pathways that serve to guide action; in the design solution such an organism implements, the content of these signals is *practical*.[46] Humans are equipped to inhabit much more variable environments than some other species, and so the contents of those signals are correspondingly flexible: some of them mark actions one has performed as successes, and others as failures; some mark changes in circumstances as improvements, and others as deterioration; some mark attainments or acquisitions as desirable, and others as undesirable; some mark avenues of exploration as promising, or the contrary. These signals are functionally, in all respects, practical observations; in humans, they are not used raw, but typically serve as inputs to inference and practical theorizing. A philosopher's example: one might experience a series of straight readings of Nietzsche as disappointing, and conclude that one should not spend any more time on them, but look instead to other ways of reading Nietzsche.[47]

Moral philosophers who have been around the block a few times will be quick to leap to conclusions: that I must be endorsing a form of 'moral realism,' on which we perceive—using sense modalities unheard of by science—what we might as well call *helicopter values*. (Like helicopter parents, they *hover*—in this case, over the physically existing objects in our world.) So notice that the signaling pathways I am gesturing at include the responses currently being investigated by hedonic psychologists. And while this is not the place to lay out a metaethical position, please do not assume that I hold moral realist views; on the contrary.[48]

[46]For discussion of such signals in some nonhuman organisms, see Sterelny, 2003.

[47]By a 'straight reading,' I mean one that treats Nietzsche as a philosopher engaged in the first place in producing a theory that we aim to reconstruct, and that ignores or only cursorily acknowledges the elephant in the room, that Nietzsche *writes funny*. My own test for whether my students see them as disappointing illustrates how we exploit the signaling systems we have available. After presenting someone's straight reading, I have them vote on whether an academic now advancing this view would be worth their attention. After a few rounds of this, the students generalize from their own votes, and conclude that straight readings, of this particular philosopher, are not worth their attention. Since the students have almost uniformly been brought up to think of the history of philosophy as the pursuit of such readings, we have a slightly exotic example of practical observation correcting policies in unplanned-for ways; this is autonomy at work in philosophy.

[48]For more of the picture, and supporting argument, see Millgram, 1997, Millgram, 2000, Millgram, 2004 (and since I have had readers take *Practical Induction* for a moral realist tract, a belated reading instruction: it is carefully phrased throughout to avoid any such commitments). It does seem to me that, like other entries that make up the

Serial hyperspecializers are explorers of their environment, and their philosophers owe them an understanding of self-government on which it makes sense of autonomous exploration. When someone goes looking for a shortcut to Asia, and instead discovers America, or embarks on a surveying and bartering expedition, and instead conquers an empire, previous plans have become irrelevant, and a successful response (whether by that person or his successors in the field) consists in large part in being willing to cast off the previous standards that had guided the enterprise, and to adopt novel reconceptions of success—and with them, new hopes and ambitions—rendered appropriate by serendipity.[49] Our schoolchildren are taught to admire the heros of the age of exploration as models of autonomous agency, and for once the schoolchildren are being taught properly. Our philosophical theorizing should at least live up to what every schoolchild knows.

8

Recall that Bratman's Master Move appealed to what is the mainstream view of personal identity among analytic philosophers today: what makes you the same person you used to be is psychological continuity, typically glossed as remembering your past, acting on your former intentions, having a similar character, and so on. However, it should be obvious that identity concepts need to be tailored to species.[50] A butterfly is not psychologically

---

standard menu of metaethical positions today, moral realism is a metaethics unsuited to serial hyperspecializers. That's a promissory note, but for some of the preparatory work, see Millgram, 2010.

[49]The first of those cases shows that the adjustments can proceed quite unevenly. Columbus never abandoned the conviction that he had reached the Indies, and almost until the end of his life continued to look for access to China and Japan; this despite his encountering populations that were not plausibly the periphery of the advanced manufacturing and trading economies that he sought. On the other hand, he very rapidly formulated and implemented colonial policies that—however repugnant—served as the basis for the Spanish colonial enterprise: policies that were effective against the background of the cultures he did encounter, and which it would have made no sense to adopt in the immediate vicinity of a great Asian power. For an overview, see Morison, 1992, esp. pp. 278f, 290f, 355f, 380f, 464–67, 553f.

Prescott, 2000, p. 137, summarizes the initial scope of the mandate given to Cortés, and pp. 614f look back on one of history's most astounding cases of mission creep.

We should not, however, assume that exploration is always well-managed; again, it is not as though whatever cognitive mechanisms we have to guide us are infallible. In the literature of the age of exploration, *La relación* (Cabeza de Vaca, 1542/2002), is a suitable corrective.

[50]This is a Wiggins-like observation, but I wouldn't want to buy into all of the details of his treatment. For the latest revision, see Wiggins, 2001.

continuous with the caterpiller it was. But it is still the same animal, because metamorphoses are what butterflies *do*: that's their *ergon*, or, in a more modern locution, metamorphoses are part of the design solution they implement.

We metamorphose psychologically: that's what it *is* to be a serial hyperspecializer and a segmented agent. In moving from niche to niche, you can quite correctly throw out your old goals, standards, preferences, intentions and policies wholesale; you can also, quite correctly, throw out almost all of your memories. When this happens, you are still someone in whom your former self has a prudential stake (this being what really matters, in this philosophical subject area, about being the same person)—at any rate, provided the transition to your new psychology was managed using the modes of rationality appropriate to segmented agency. For instance, when cognitive signals such as boredom and frustration prompt you to move on to a new niche, to forget your former life as thoroughly as possible, and to take up entirely new activities that you can find interesting, and in which you feel yourself competent and at home, that is a benefit to *you*.[51]

Neo-Lockean accounts of personal identity are just about right for a creature that, on the one hand, is like us in having a mental life, but on the other, is designed (or rather, 'designed') to live out its life in a single, stable niche. Such a creature can be imagined as starting life with a program that

[51] Instrumentalist theories of practical rationality are unable to account for this truism; see Williams, 1973, ch. 6.

These days, sophisticated neo-Lockean accounts take time out to define psychological continuity in such a way that there may in fact be very little psychologically in common between two temporally widely separated stages of a person: you will count as psychologically continuous with your future self if all of the adjacent pairs of intermediate stages share *some* of their psychology, and the relation that either is taken for or taken to replace personal identity is then introduced as the ancestral of that relation between adjacent stages. (Put more concretely: you may remember nothing of your five-year-old self, but you are still that person because you remember some of your ten-year-old experiences, and your ten-year-old self remembered some of his five-year-old experiences; see, for instance, Parfit, 1987, pp. 205f.) That much psychological continuity will often be found in segmented agents; for instance, they may remember why they gave up on a previous niche.

But how is one to motivate the psychological continuity approach? Surely by way of the thought that the degree of connectedness is what matters, and even if it's not *always* true that the more in the way of connections, the better, what matters in neo-Lockean personal identity would be getting short shrift in a human animal whose stages (*properly*, on some occasions) shared only a tiny handful of psychological states with their predecessor and successor stages. For segmented agents, it will often be the case that the *less* in the way of continuity, the better. The more efficient your garbage collection algorithms, the more effectively you can devote your cognitive resources to mastering your new niche.

will govern its activities until the end of its days—a design approach that could be effective within the confines of a stable niche. The psychological states that the program deposits and uses—records in memory of the creature's progress and the state of its environment, overarching goals, subgoals and the like—may (and ought to be) updated constantly. But they will never need to be deleted wholesale, and if they *are* deleted wholesale, an individual creature of this kind will not normally survive the operation. So treating *such* a creature's life as coextensive with a continuous psychology of this kind is entirely reasonable. We are not such creatures; a neo-Lockean account of personal identity is inappropriate for us.[52]

Return to the rough-but-widely-shared picture of unified agency with which we began. That's a pretty good picture of a creature for whom a neo-Lockean theory of personal identity would be appropriate. But it's not at

[52]A less popular but still respectable position on personal identity prefers bodily continuity as the criterion of sameness—you are the very same person you once were if you have the same body—and it might seem that my account is committed to this alternative. I am not at all certain, for two reasons. First of all, I take seriously Bernard Williams's arguments to the effect that we do not have a philosophically satisfactory account of the body, and that we have not thought through what our distinction between body and mind comes to (1973, chs. 1–5, and esp. pp. 11f, 68ff)—though, oddly enough, Williams himself went on to endorse a bodily-continuity account of identity, which strikes me as an uncharacteristic failure of philosophical nerve. 'Body' (and, more recently, 'organism'), in these discussions, is just a placeholder, a we-know-not-what. Second, we do not have an explanation for bodily continuity being the basis of an identity concept suitable for segmented agents, and, without that explanation, we should not just accept whatever looks like the leftover theory.

The other side of a philosophical theory of personal identity is a philosophical theory of death; after all, you are dead once there is no one who is identical to you. (For one expression of this insight, see Parfit, 1987, pp. 281f.) I have been gesturing at an account of what it is to be a human being, one on which the fact that you are going to die is not an essential part of the design, but rather on the order of a manufacturing flaw. Death is a form of planned obsolescence suitable for creatures that—conformably to the mainstream model of agency—cannot reprogram themselves to adjust to deeply different environments. When a creature is self-reprogramming, throwing out the hardware platform at one-generation intervals looks *wasteful*.

That might suggest that something is deeply amiss in the account. (I'm grateful to Havi Carel for pressing me on this point.) It is a widely held view that your death is the frame in which the elements of your life are meaningful, and that to lose sight of your death is to live 'inauthentically'. A suspicious reader might even wonder whether the present account is not just a way of avoiding the confrontation with one's own mortality. Now, certainly there is no point in pretending that one is not going to die. But in the segmented form of agency at which I have been gesturing, the frames that make activity meaningful may well be much shorter than an entire life. To face up to one's death does not mean: to think about it in terms that would be suitable only for the much simpler sort of animal implicitly presupposed by mainstream theory of agency.

all a satisfactory rendering of a segmented agent. Segmented agents are not completely fragmented; we will see a great deal of top-down, policy-governed coordination of activity within each segment. But we should expect to see substantial agential disunity induced by fault lines between niches, and by the cognitive devices that facilitate niche exploration and niche jumping.[53] Like theorizing about personal identity, recent theoretical work on agency has been rich, subtle, and interestingly argued, but also philosophizing suited to a species not our own.

That disunity will be both diachronic and synchronic. I mentioned in passing that segmented agents will not infrequently pursue activities that go on in different niches. (The picture of segmented agency, recall, was a first approximation, and this is one important qualification to add to it.) Because the evaluative maps of these niches will typically amount to incommensurable systems of standards, the various activities in which segmented agents engage will often fail to be governed by a unified hierarchy of evaluative judgments and goals. They may even be at cross-purposes, and although a segmented agent is to some degree equipped to manage resource competition between projects and activities, we should not expect these to be regulated by higher level goals, desires or standards. That would be to commit the error of thinking that an engineering solution which works within niches can be scaled up to solve problems that span niches, as though the world were simply a larger niche, and could be navigated by constructing a system of standards—such as a policy—suitable to a much more constrained environment, only *bigger*.

I suggested that segmented agents must be cognitively equipped with prompts that tell them when it is time to switch niches, and that frustration and boredom are probably representative prompts of this sort. Let's conclude by considering what this means for the familiar philosophical interest in superlative attribution. When they respond to such prompts, segmented agents can surprise themselves by taking steps that contravene their current projects: steps they do not endorse, steps that do not ensue on their policies for weighing reasons, steps that lead them to say, "I didn't really choose

[53]Sometimes readers are reminded of the view advanced by Strawson, 2009, on which 'selves' are also much briefer than human lives, so it's worth pointing out that my own view and his have very little in common. Strawson's 'selves' are momentary, understood to be purely mental, i.e., distinct from the bodies with which they are associated, and the best example we have of objects; they have nothing to do with specialization, and any animal that has experiences of any kind is supposed to have them. The agential segments I am exploring are rarely of shorter duration than several years, are not supposed to be in any way disembodied, are not being advanced as being 'objects' or substances, and are side effects of an ecological strategy of specialization to available niches in one's environment.

to do it... it just kind of happened." When they act, they may not know what they are doing, or why. ("I don't know why I did that: I just *found* myself doing it. I really surprised myself.") And when they behave in this way, they may well be evincing, not their irrationality, and not their lack of self-government, as the theoretical mainstream has it, but what is precisely practical rationality and autonomy for creatures of this kind.

Mainstream theory of agency is very good at articulating the phenomenology of acting on the basis of an override—a signal that is independent of your system of standards, of your complex of goals and desires (in one famous bit of terminology, of your 'subjective motivational set'), of your practical identity, and so on. (The accompanying utterances at which I just gestured are typical, and it can feel as though it wasn't *you* who did it. Although of course you can often *explain* why you did it; for instance, you were bored out of your mind.) Because mainstream theorists identify the disposable personae you happen to be projecting at the moment with your *self*, they treat your responding to such signals as a disaster: as the unraveling of your agency, and thus of your personhood. But if you are a segmented agent, it is not necessarily a disaster at all; this is just how you cast off devices that are suitable, temporarily, for coping with one environment, in favor of other devices that are suitable for other environments. If you like, and granting, just for a moment and for the sake of the argument, that the mainstream is giving a successful analysis of what it is to be an agent, you are not a single agent, but rather the substrate of a series of them. Agents are interfaces you conjure up to meet the needs of the moment. Do not make the mistake of thinking that one or another of them is *you*.

References

Bratman, M., 1987. *Intention, Plans and Practical Reason.* Harvard University Press, Cambridge, Mass.

Bratman, M., 1999. *Faces of Intention.* Cambridge University Press, Cambridge.

Bratman, M., 2001. Taking plans seriously. In Millgram, E., editor, *Varieties of Practical Reasoning*, MIT Press, Cambridge, Mass.

Bratman, M., 2006. *Structures of Agency.* Oxford University Press, Oxford.

Bratman, M., 2009. Intention, practical rationality, and self-governance. *Ethics*, *119*, 411–443.

Cabeza de Vaca, A. N., 1542/2002. *Chronicle of the Narváez Expedition*. Penguin, New York. Translated by Fanny Bandelier, revised by Harold Augenbraum.

Chesnow, Z., 2008. *Kol ha-Neharot Zormim la-Yarden*. Philobiblion, Jerusalem.

Christman, J., editor, 1989. *The Inner Citadel*. Oxford University Press, New York.

Dworkin, G., 1988. *The Theory and Practice of Autonomy*. Cambridge University Press, New York.

Enoch, D., 2006. Agency, schmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review*, *115(2)*, 169–198.

Ferrero, L., 2009. Constitutivism and the inescapability of agency. *Oxford Studies in Metaethics*, *4*, 303–333.

Ferrero, L., 2010. Decisions, diachronic autonomy, and the division of deliberative labor. *Philosophers' Imprint*, *10(2)*, 1–23.

Fleischacker, S., 1999. *A Third Concept of Liberty*. Princeton University Press, Princeton.

Foucault, M., 1984. What is an author? In Rabinow, P., editor, *The Foucault Reader*, pages 101–120, Pantheon, New York.

Frankfurt, H., 1988. *The Importance of What We Care About*. Cambridge University Press, Cambridge.

Frankfurt, H., 1999. *Necessity, Volition, and Love*. Cambridge University Press, Cambridge.

Friedman, M., 2001. *Dynamics of Reason*. CSLI Publications, Stanford.

Hill, T., 1991. *Autonomy and Self-Respect*. Cambridge University Press, New York.

Kavka, G., 1983. The toxin puzzle. *Analysis*, *43*, 33–36.

Korsgaard, C., 1996a. *Creating the Kingdom of Ends*. Cambridge University Press, New York.

Korsgaard, C., 1996b. *The Sources of Normativity*. Cambridge University Press, Cambridge.

Korsgaard, C., 2008. *The Constitution of Agency*. Oxford University Press, Oxford.

Korsgaard, C., 2009. *Self-Constitution*. Oxford University Press, New York.

Kuhn, T., 1970. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.

Martin, S., 1998. Changes in the memory after fifty. In *Pure Drivel*, pages 18–21, Hyperion, New York.

McDowell, J., 1998. *Mind, Value, and Reality*. Harvard University Press, Cambridge, Mass.

Millgram, E., 1997. *Practical Induction*. Harvard University Press, Cambridge, Mass.

Millgram, E., 2000. What's the use of utility? *Philosophy and Public Affairs*, *29(2)*, 113–136.

Millgram, E., 2004. On being bored out of your mind. *Proceedings of the Aristotelian Society*, *104(2)*, 163–184.

Millgram, E., 2005. Practical reason and the structure of actions. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Millgram, E., 2009a. Applied ethics, moral skepticism and reasons with expiration dates. In Black, S. and Tiffany, E., editors, *Reasons to Be Moral Revisited*, University of Calgary Press, Calgary.

Millgram, E., 2009b. D'où venons-nous... Que sommes nous... Où allons-nous? In Callcut, D., editor, *Reading Bernard Williams*, Routledge, London.

Millgram, E., 2009c. Practical reasoning for serial hyperspecializers. *Philosophical Explorations*, *12(3)*, 261–278.

Millgram, E., 2010. Mètaètica mìlliana. *Quaderni di Scienza Politica*, *17(2)*, 245–265.

Morison, S. E., 1992. *Admiral of the Ocean Sea: A Life of Christopher Columbus*. Book-of-the-Month Club, New York.

Nagel, T., 1978. *The Possibility of Altruism*. Princeton University Press, Princeton.

Nozick, R., 1981. *Philosophical Explanations*. Harvard University Press, Cambridge, Mass.

Parfit, D., 1987. *Reasons and Persons*. Oxford University Press, Oxford.

Prescott, W., 2000. *History of the Conquest of Mexico and History of the Conquest of Peru*. Cooper Square Press, New York.

Rose, P. L., 1990. *German Question/Jewish Question*. Princeton University Press, Princeton.

Sterelny, K., 2003. *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell, Oxford.

Strawson, G., 2009. *Selves*. Oxford University Press, Oxford.

Tubert, A., 2010. Constitutive arguments. *Philosophy Compass*, *5*, 656–666.

Velleman, J. D., 2000. *The Possibility of Practical Reason*. Oxford University Press, Oxford.

Vogler, C., 2002. *Reasonably Vicious*. Harvard University Press, Cambridge, Mass.

Watson, G., 2004. *Agency and Answerability*. Oxford University Press, Oxford.

Wiggins, D., 2001. *Sameness and Substance Renewed*. Cambridge University Press, Cambridge.

Williams, B., 1973. *Problems of the Self*. Cambridge University Press, Cambridge.

Williams, B., 1981. Moral luck. In *Moral Luck*, Cambridge University Press, Cambridge.