

DESIRES...AND BELIEFS...OF ONE'S OWN¹

Geoffrey Sayre-McCord and Michael Smith

Much work in recent moral psychology attempts to spell out what it is for a desire to be *an agent's own*, or, as it is often put, what it means for an agent to *identify with* certain of her desires rather than others.² The aim of such work varies. Some suggest that an account of what it is for a desire to be an agent's own tells us what it is for an agent to be free or autonomous.³ Others suggest that an account of what it is for a desire to be an agent's own tells us what it is for an agent to value something.⁴

According to the most influential account of what it is for a desire to be an agent's own developed by Harry Frankfurt, a desire is an agent's own if she has a higher order desire concerning it.⁵ She must desire not just that she has that desire, but that that desire leads all the way to action, and the higher order desire in turn must be one with which she is satisfied, in the sense of being one to which she does not have, at some level, an even higher order aversion.⁶ With this account on hand, the suggestions concerning freedom or autonomy, on the one hand, and valuing on the other, are as follows. An agent acts freely, or autonomously, if the desires that enjoy higher order ratification, rather than others, are effective in action, and an agent values the object of any desire that enjoys such higher order ratification. Though this hierarchical account of what it is for a desire to be an agent's own has many supporters, we think it faces at least two problems, problems that inspire us to look for an alternative.

The first problem is that the account does not readily generalize. It makes sense, or so we think, to ask not just whether a certain desire is an agent's own, but also whether a certain belief, or more generally a view, is an agent's own. Agents can, in other words, identify more or less not

¹ We are grateful for feedback from audiences at the University of North Carolina at Chapel Hill (where Michael Smith delivered the paper) and the Australian National University (where Geoff Sayre-McCord delivered it).

² The *locus classicus* is Harry Frankfurt's "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* (68) 1971, pp. 5-20. Frankfurt further developed the idea in "Identification and Wholeheartedness" in Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press, 1987).

³ This was Frankfurt's aim.

⁴ This seems to have been at least part of what Gary Watson was doing in his "Free Agency," *Journal of Philosophy* (72) 1975 pp.205-220. Michael Bratman develops both ideas in various places, including "Valuing and the Will", *Noûs* 34 (14) pp.249-265; "A Desire of One's Own," *Journal of Philosophy*, May 2003, pp. 221-242; and "[Anchors for Deliberation](#)" in Christoph Lumer and Sandro Nannini (eds.), *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy* (Farnham, Ashgate Publishing, 2007). Frankfurt argues that we cannot use the concept of a desire of one's own to illuminate the concept of what it is to value something in his contributions to Sarah Buss Lee and Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge, Mass.: MIT Press, 2002).

⁵ "Freedom of the Will"

⁶ "Identification and Wholeheartedness"

just with their desires, but also with their beliefs.⁷ But it is unclear how we might generalize the hierarchical account of what it is for a desire to be an agent's own so as to turn it into an account of what it is for a view or a belief to be an agent's own. It seems unpromising, for example, to suppose that a belief is an agent's own just in case she has a higher order desire with a similar content concerning it: a desire to have the belief and for that belief to lead all the way to action. That sounds more like wishful thinking than identification.⁸ Nor is it promising to suppose that a belief being the object of a higher-order belief with a similar content makes the crucial difference; that a belief is an agent's own if she has a belief that she has that belief and that it leads all the way to action. The beliefs with which agents identify may have to be ones the agent has beliefs concerning, but we doubt that they need to lead all the way to action. Moreover, as we will argue below, a person may properly be identified with beliefs she does not realize she has.⁹

The second problem with the hierarchical account is that it does not allow us to distinguish properly, in a way we think people can and do, between identifying which desires (and beliefs) are actually our own and identifying which desires (and beliefs) should be our own if we are to be free and autonomous. Securing such a distinction is crucial if we are to acknowledge that central aspects of who we are, aspects that shape what and why we do what we do, may nonetheless be features of ourselves of which we disapprove. One cannot always disown what one wishes one could. That an alcoholic might sincerely desire that her desire to get drunk not be effective in action is perfectly compatible with her first order desire to drink being a central part of who she is, and something that she could deny or disown only disingenuously. An account of which desires (and beliefs) are ones own should leave room for this fact.

Our aim in this paper is thus to develop an alternative and more general account of what it is for both desires and beliefs to be an agent's own, an account which allows us to distinguish quite sharply between the descriptive and the normative questions. Although our paper might be read as an attack on the hierarchical account of what it is for a desire to be an agent's own, we do not intend it that way. For all that we say here, the hierarchical account might well provide us with *a sense* in which desires are an agent's own, and hence with *a* concept of identification. In particular, we are thinking it might plausibly capture an account of when an agent willingly owns particular desires (and also, we suspect, beliefs) and so willingly owns the actions to which these attitudes give rise. But what one willingly owns is importantly different from what is recognizably one's own in a different sense—what one is accurately identified with, either by oneself or others. And it is different too, we think, from what makes for autonomy.

The crucial question, as we see things, is not whether there is a sense of a desire being an agent's own, or identification with one's desires, that is accurately captured by the hierarchical account, but rather whether there is important work to be done that calls for a different account of

⁷ It makes sense too, we suspect, to talk of emotions, moods, and actions as an agent's own, or as aspects of themselves with which they may be identified. It is unclear how we might generalize the Frankfurtian account to these as well.

⁸ In fact, there are two ways to extend the Frankfurtian account to belief. One is the suggestion made in the text that a belief is an agent's own just in case she desires to have the belief, the other is that the relevant desire is a desire that the belief be effective in action. In neither case is the resulting view very plausible.

⁹ The view we go on to defend does, however, assign a crucial role to a higher-order belief.

identification, of which desires and beliefs are one's own, and of what counts as having been done autonomously. We think that there is. The different account we develop makes sense of people accurately recognizing themselves as, actually, different than they would want to be, without treating their shortcomings as alien or as not really who they are. Moreover, we think, it makes better sense both of what it is for an agent to be autonomous, and of what valuing is. We will make some all too brief comments about autonomy and valuing at the very end.

Before moving to our account, we want to say a little about what we take to be the most plausible version of the hierarchical account, that offered by Michael Bratman.¹⁰ On Frankfurt's proposal, recall, a desire counts as one's own only if one has a higher-order desire that the desire in question be effective in action. As Frankfurt himself recognized, it seems not just any old higher-order desire of this sort will be enough. If, for instance, the higher-order desire is one concerning which one has an even higher-order desire that one not have it, or that it not be effective in action, the original first-order desire seems still not to be one's own. (It would certainly be strange to see it as one with which one willingly identifies.) Frankfurt's response to this concern is to require that the ownership-inducing desire be free of any higher-order rejection. But that has struck many as not enough. Even if a higher-order desire stands unchallenged at a still higher level, it seems that it might be either unacceptably arbitrary or independent of one's values in a way that invites the idea that the desires it ratifies have no special standing as what one willingly owns (as opposed, say, to merely finding oneself with). Pursuing this idea, Gary Watson suggested that the relevant higher order states must express, or reflect, one's values if they are to constitute the lower order desire as one's own.¹¹ One is acting freely, and on desires that are one's own, he argues, when they reflect and are aligned with one's judgments of what is worth doing.

Bratman's proposal follows naturally on this idea, but resists appealing to judgments of what is worth doing. Instead, he suggests, a desire counts as one's own when one has a higher-order intention or policy of treating that desire as reason-providing. Bratman's idea is that desires that enjoy this role have a privileged place in agency and gives them a kind of authority for the agent that means they, and the actions to which they lead, are rightly counted as the agent's own. "These intentions and policies," he explains, "involve distinctive commitments concerning associated forms of practical thought and action, and play central roles in the cross-temporal organization of our temporally extended lives."¹² But they are not themselves judgments to the effect that the desires in question are reason-providing, nor (it seems) need they reflect judgments that the intentions or policies are worth having.¹³ While Bratman does not consider the questions of whether beliefs might count, or not, as one's own, we think his proposal does plausibly generalize to beliefs, recommending the view that a belief counts as one's own if one has a higher-order policy of treating the content of that belief as reason-providing. (And,

¹⁰ "A Desire of One's Own"

¹¹ "Free Agency"

¹² "Free Agency" p. 222

¹³ There are of course interesting complications involved in characterizing what is involved in treating something as reason-providing. And it is important to stay clear on the difference between treating the object of a desire as reason-providing and treating the desire, or one's having the desire, as reason-providing. For our purposes, though, these complications and details are not important to sort through.

presumably, it generalizes as well to other attitudes as long as they are ones towards which one might have such a policy.)

Bratman's proposal differs in significant ways from Frankfurt's, not least in its appeal to the normative notion of being reason-providing. So it is worth noting that many are attracted to Frankfurt's account in the first place because it eschews any appeal whatsoever to normative concepts in giving an account for what it is to think of something as valuable and what it is for a desire to be one's own. However, we have no worries about introducing the notion, and in fact think that something like it is essential if we are to capture accurately the idea of autonomy. Bratman's proposal differs significantly from Watson's as well, it seems, in that having the sort of intention or policy Bratman identifies is (as we understand it) a matter of having certain commitments, dispositions, and modes of practical thinking that don't themselves involve forming any beliefs about what is reason-providing. In fact, we gather, an agent can have the requisite intention or policy, and act on the basis of desires it *treats as* reason-providing, without thinking they are reason-providing.

As we see things, the difference between Bratman and Watson leaves in place the concern that emerges when one asks whether an agent can have such a policy and yet believe of the desire (or the belief) in question that it is not actually reason-providing. If so, we suspect, having the status of an as-if reason-providing consideration does not plausibly work to privilege a desire as, in any important sense, one's own. If not, if no such belief is compatible with having the required intention or policy, our hunch is that that is because having the intention or policy is tantamount to thinking the desire is reason-providing. It's not that we would have a complaint about that—in fact, at least when it comes to autonomy, we think having some such belief is necessary. But we do suspect that Bratman is trying to offer a higher-order attitude that does not constitute, nor entail, the sort of evaluative judgment Watson argued, and we agree, was necessary.

In fact, our understanding is that Bratman puts no real normative constraints on what an agent might have a plan to treat as reason-providing (other than a few quasi-structural norms that might restrict what combination of desires might be taken as reason-providing). We think, in contrast, that when it comes to understanding the nature of freedom and autonomy, normative constraints are crucial. At the same time, however, we will be arguing that there is an important sense in which a belief or a desire is one's own—one with which a person properly identifies, or is identified by others—that is independent both of what normative considerations are in play and of what the agent might think, or be committed to treating as, being in play.

Since our aim is to give a quite general account of what it might mean for desires and beliefs to be an agent's own, our strategy will be to focus on key structural features that beliefs and desires have in common. Specifically, we maintain that desires and beliefs alike come in degrees along two dimensions and that one dimension—a dimension that we will characterize as *relative stability*—is central to understanding which beliefs and desires are an agent's own. We begin by focusing on the case of belief, and then proceed to consider the case of desire.

1. Beliefs of one's own

Note that beliefs quite generally have the following two features. There is the degree of belief that agents have in the propositions they believe, and there is how stable the degree of belief is

under the impact of experience, new information, and reflection. We will consider these two features in turn.

The first feature we mentioned is the degree of belief an agent has in the propositions she believes. For example, someone might be highly dubious that the sun will explode tomorrow, but confident that it will rain tomorrow, and even more confident still that there will be a football match tomorrow.¹⁴ This difference in her degrees of belief is the sort of thing that gets revealed in how much she would be willing to bet on one outcome as opposed to another under circumstances of forced choice. It is a *synchronic* fact about her beliefs.

The second feature is quite different, in this regard. The second feature concerns how stable the degrees of belief are, over time, in the light of experience, incoming information, and reflection. For example, though someone might believe to the very same degree the proposition that the Sydney Swans is the strongest Australian Rules football team and the proposition that her son is a responsible supermarket employee, her degree of belief in the latter might be a very stable feature of her psychology, whereas her degree of belief about the Sydney Swans might be notably less stable. While virtually nothing in the way of experience or incoming information would change her degree of belief in the proposition that her son is a responsible supermarket employee—she will remain confident about that come what may—all sorts of incoming information would radically change her degree of belief in the proposition that the Sydney Swans is the strongest Australian Rules football team.

As we see it, the stability at issue may be merely a matter of actual survival, or it may be a matter of counterfactual survival. For different purposes, different measures of stability seem to be more or less relevant. The first would seem to travel naturally with identifying identification with a descriptive characterization of *de facto* psychology; the second, which isolates *resistance* to change, would seem to travel more naturally with something like centrality to character, and so with thoughts about what is characteristic of the person. Though we suspect the two interact, the second is ultimately the notion that we will lean on in our account of identification.

Let's call beliefs that are stable across time 'robust', and those that are not stable 'fragile'. Just how robust or fragile a particular belief might be is, clearly, a matter of degree, and beliefs might be robust or fragile for a whole range of quite different reasons, some of which reflect the agent's rationality, and some of which do not. Crucially, though, no matter how robust or fragile an agent's beliefs might be, and no matter what explains their being robust or fragile to the degree that they are, facts about the robustness and fragility of an agent's beliefs are plainly fixed *diachronically*. It is this diachronic feature of beliefs that will be important in what follows.

As we said, an agent's beliefs might be robust or fragile for a whole range of quite different reasons. Focusing first on robust beliefs, and looking at one extreme, a belief might be one which an agent has stably over time because she has canvassed for relevant evidence thoroughly, thought about the matter carefully, and, as a result, can see a whole host of interconnected reasons for thinking that the proposition she believes is true. In other words, the belief may be

¹⁴ Of course, a person's beliefs might themselves be about objective probabilities, e.g. about the likelihood that a particle will leave a trace in some experiment, so it is worth noting that these beliefs too will admit of different degrees of belief in just the way that other beliefs do.

robust because it is, given the information available to her, evidentially extremely well supported, so well supported that rationally rejecting it would involve abandoning, or at least altering, a host of her other beliefs.

At the other end of the spectrum, among robust beliefs, an agent's belief might be stable because she is irrationally disposed to cling tightly to it. The belief might be the product of wishful thinking, say, so that no matter what countervailing information comes in, the agent will be disposed to reinterpret that information, or ignore it, or the information will in some other way be prevented from having its proper evidential impact. The mother who is confident that her son is a responsible supermarket employee, as we are imagining it, is a case in point. She clings to the belief no matter what the incoming evidence because the belief isn't based on the evidence at all.

There will be cases in between as well. A belief might be robust not because locked in by wishful thinking, and not because the agent has thought about it and seen so many reasons for assigning the proposition believed the degree of belief she assigns it, but because, though she has thought about the matter, she is not very good at seeing the implications of her evidence, or because she is vulnerable to certain common biases, or because she fails to think things through carefully.

The same point applies to fragile beliefs. Some beliefs are fragile because agents are in situations in which evidence is sparse or shifting, and an agent's changes in belief simply reflect the vigilant exercise of her capacity for rational belief revision. The mother who believes that the Sydney Swans is the strongest Australian Rules football team might be in this situation. Her belief is fragile because the evidential situation, to which she is highly attuned, is constantly changing. At the other extreme, some agents' beliefs are fragile because the agents in question forget what it was that they came to believe in the first place, or because they are prone to make all sorts of errors in their assessment of the incoming evidence, or because they are systematically influenced by what those who are around them think, where those who are around them vary from time to time as regards what they think.

With this distinction between robust and fragile beliefs in place, we can now offer what seems to us to be a clear and intuitive suggestion about what it might mean for a belief to be an agent's own: that is, for a belief to be one with which she is properly identified, and with which she could accurately identify. Our suggestion, perhaps unsurprisingly, is that a belief is an agent's own, or one with which she is to be identified and with which she could accurately identify, if it is robust: that is, stable in a way that qualifies it as characteristic of her. And a belief will then be one with which an agent does in fact accurately identify, in the sense we are after, if she recognizes it as one with which she is to be identified: that is, if she knows that it is characteristic of her. A belief that is fragile, by contrast, is one which is not an agent's own, not one with which she is to be identified, and not one with which she could accurately identify. The appeal of this suggestion is, we hope, plain, but, in case it isn't, we offer the following by way of support.

The robust beliefs that agents have, no matter how their robustness is explained, constitute the distinctive world-view that those agents have. The point is perhaps most obvious when the view has the sort of wide scope that religious and political doctrines often do. In these cases, when an agent's allegiance to such a view has proven robust—when her belief in Catholicism, or

Communism, or Capitalism, has truly taken hold—the agent and others properly come to identify her with that view. But the same point emerges nicely, if less dramatically, with beliefs having a much narrower scope. Consider a person's beliefs about her own past. Here too one comes to identify oneself in terms of the events concerning which one has robust beliefs, as do others. The memories rehearsed, the stories reviewed, and the events relived, especially in the company of intimates, serve to constitute our shared sense of our selves and of who we are.

It is worth noting that this is true of such beliefs whether or not they are beliefs we desire to have, or to have effective in action. No small proportion of the beliefs with which agents are properly identified, and with which, often, agents do identify, are beliefs that they themselves would rather be without, often for good reason, and many of them are simply irrelevant to action, except of course to the acts of speech in which we give them expression. Nonetheless, to the extent the beliefs prove to be robust they will, in an important sense, emerge as the agent's own and as beliefs with which she and others can and will properly identify her.

By contrast, the fragile beliefs that agents have, precisely because they are fragile, cannot, in this sense, be said to be the agents' own, and nor are they ones with which she can reasonably be identified. They do not help to constitute the agents' distinctive perspective on the world. Consider those whose beliefs shift as fashion does, or those who are constantly reinventing their past, or those whose memories wax and wane. In all these cases, the agents in question might well come to be identified with their very variability—that, after all, is by assumption a robust characteristic they exhibit—but none of their beliefs, no matter how fervently held at one point or another, will be something with which they will properly be identified.¹⁵

Note that our account leaves open the possibility that an agent might be identifiable with a belief—that is, a belief may be characteristic of her—even though she does not herself believe that it is characteristic of her, and hence even though, in our sense, she does not identify with it. (Here we are supposing a person identifies with a desire or a belief if, but only if, she believes of it that it is one with which she is accurately identified.) Sexist and racist beliefs are, we think, very often like this. Those who have such beliefs will, when pressed, try to explain why the sexist and racist things that they say and do aren't really evidence of their holding sexist and racist beliefs. But though that suffices to show that they do not identify their sexist and racist beliefs as their own, the sexist and racist desires and beliefs may nonetheless still (in our sense) be ones with which they are properly identified. These desires and beliefs might well still constitute (a disturbing aspect of) their distinctive perspective on the world. The fact that they do not recognize this is neither here nor there.

Similarly, our account allows for the possibility that an agent may identify with a belief—that is, believe that it is characteristic of her—and yet be mistaken, either because she lacks the belief that she takes herself to have, or because, though she has the belief, it is not sufficiently robust to count as one with which she is to be identified. Thus, to go back to those who are sexist and racist, they may well believe that they did, do, and always will believe in equality come what

¹⁵ This is, we think, a slight exaggeration, as an agent might well come to be identified with a really bizarre (or otherwise extremely memorable) belief, even if the belief is fragile. But what is true even in these cases, we suspect, is that the fragile belief comes to be identified with the agent thanks to the mediation of a robust belief that she once held the bizarre view with which she has come to be identified.

may, and yet be wrong. And it is not unusual to meet enthusiasts who bounce from fad to fad, professing that this time there will be no change of mind as they've finally found what they've been looking for all along, without any of the beliefs that they thus acquire being stable enough to count as ones with which they are properly identified. It seems to us that mistakes of both these kinds are not only possible but, when they occur, revelatory of how well-attuned we are to recognize human frailty. We have little difficulty writing off the rationalizations of sexists and racists as mere noise in the system, nor are we long tempted to take seriously the professions of conviction made by those whose enthusiasms are constantly changing.

Robust beliefs constitute an agent's distinctive perspective on the world, no matter how their robustness is to be explained. And one of our fundamental interests in other people lies in figuring out what their distinctive perspective on the world is. Still, it is important to remember that we often care about more than just what their perspective on the world happens to be. In many cases we also care about what explains the robustness of the beliefs that constitute their perspective. The example of those with sexist and racist beliefs who disavow their beliefs is a case in point. We all have an interest in knowing when it is worth engaging people in a conversation about what they believe and when it would be pointless to do so. But this isn't the only example.

Consider the broad range of areas in which we seek out experts in particular fields. For example, think about the way you go about choosing where to get your car repaired, or which recipe book to buy, or whether to believe in climate change. Experts count as experts, and are identified with their expertise, only because their beliefs in the relevant area have become robust. Of course, mere stability in the face of evidence and reflection does not an expert make. Those whose beliefs are stable but consistently wrong, or stable but quickly out of sync with the latest thinking, will acquire reputations for pretending to an expertise that they lack. Incoming evidence must be assessed on its merits and, where appropriate, adjustments made.

But even if someone does appropriately assess incoming evidence, she will still not count as an expert in a certain field unless her beliefs concerning the relevant matters are relatively stable in the face of her assessment of that evidence. Determining expertise in a relevant field is, we think, largely a matter of discovering whether someone's beliefs in that field are appropriately stable in the face of experience, incoming information, and reflection. A constantly shifting view disqualifies one as an expert in the relevant areas. This is not to say that the constant shifts should not be made. In many cases they should. But when they should, it is either because there is no expertise to be had or, at least, because the person whose views are shifting is not yet an expert.¹⁶

Even in areas in which we resist the idea that there are experts, part of what makes us admire certain people—part of what makes them who they are, from our collective point of view—isn't just the fact that they robustly hold certain beliefs, but is the way in which they defend their robust beliefs, resisting the various pressures they might meet to shift their views. Think of the various film reviewers, social critics, philosophers, and the like that we most admire. If they did

¹⁶ We do not mean to deny that expertise depends on much more than just having the appropriate beliefs. The skills required to translate those beliefs into action are also clearly important. Our point is merely that having only fragile beliefs, or stable beliefs that are not properly sensitive to relevant evidence, suffices to disqualify one as an expert.

not robustly hold the views they hold, and if they did not go to the lengths they do, and take the care they do, to defend their views when questioned about them, we would not admire them so much. Professional identities in such cases are, it seems to us, largely constituted by the fact that the professionals we admire robustly hold the views they hold in their area of professional expertise and have and exercise the capacity to vigorously and carefully defend those views when they are put under pressure.

To sum up our suggestion so far, the robust beliefs that agents have play a crucial role in helping constitute the perspective that those agents have on the world. Because robust beliefs play this crucial role, it seems to us that they are beliefs that can be said to be the agents' own in a clear and intuitive sense. They are part, indeed an especially important part, of who they are. As a result, they are the beliefs with which we can identify those who possess them and the beliefs with which they can identify themselves. Fragile beliefs, precisely because they do not help constitute anything by way of a perspective on the world, are not an agent's own in this clear and intuitive sense.

2. Desires of one's own

With this discussion of the case of belief in place, we now consider the case of desire. Just as we were able to distinguish two features of beliefs, so we can distinguish two similar features of desires.

First, there is the strength of the desires agents have that a certain proposition be true. For example, one might desire very strongly indeed that one's children fare well as they go through their lives, and desire as well, presumably less strongly, that one have a constant supply of good coffee. This difference in the strength of the agent's desires is the sort of thing that gets revealed in how much she would be willing to pay for one outcome as opposed to the other under circumstances where the outcomes were up to her. It is thus a synchronic fact about her desires.

Second, there is how stable the strength of an agents' desires are, over time, in light of experience, incoming information and reflection, and other pressures as well. For example, someone might have a strong desire to watch late night movies and an equally strong desire to play guitar on a regular basis—in a forced choice situation she might find it difficult to choose—but her desire that she watch late night movies might diminish, or disappear altogether, in the face of reflection and information, in a way that her desire to play the guitar on a regular basis would not.

Whereas facts about the strength of an agent's desires are fixed synchronically, and revealed by how much they would pay to have one outcome rather than another under circumstances where the outcome was up to them, facts about how stable the strength of an agent's desire is under the impact of incoming information and reflection are plainly fixed diachronically. Again, it is this diachronic feature—this time a diachronic feature of desires—that is important to our account. And it is this feature, a feature which desires share with beliefs, that allows a general account of what it is for a state to be one's own, or to be such that one identifies, or is to be identified, with it.

Let's call desires that are stable across time 'robust', and those that are not stable 'fragile'. Note that desires might be robust or fragile for a whole range of quite different reasons, some of which might reflect the agent's rationality, some of which might not. Focusing first on robust desires, and looking at one extreme, a desire might be robust because the agent has canvassed for relevant considerations thoroughly, thought about the matter carefully, and, as a result, can see a whole host of inter-connected considerations that recommend the object of her desire. It is, of course, controversial what exactly this entails, but we intend the claim to be interpreted ecumenically. If you think that the only reasons that agents can have for desiring must themselves be conditioned by other desires that the agent has, then think of an agent who only changes her desires when she believes that there is a reason to, but imagine her having a desire that coheres so well with the vast bulk of her other desires that no reason to change it exists, something that she appreciates all too well. Or if you think that there can be unconditional reasons for having desires, then think of the same case, but imagine the agent robustly believing that no considerations for getting rid of her desire obtain. In either case, what is important about these robust desires is that their stability is no sign of irrationality, quite the opposite is the case.

At the other end of the spectrum, among robust desires, an agent's desires might be stable because she has an irrational fixation. It might be a manifestation of OCD, or the product of an addiction, or a childhood trauma, or the desire might be product of wishful thinking—the agent wants to have the desire quite independently of whether there are reasons for or against it—so that no matter what countervailing information comes in, she will be disposed to reinterpret that information, or ignore it, or the information will in some other way be prevented from having its proper impact. And there are presumably cases in between as well. A desire might be robust not because it is the product of addiction or trauma, and not because the agent has reflected and come to the robust view that no reasons to change the desire exist, but rather because she has thought only a little about her desire and is vulnerable to certain systematic errors.

The same point applies to fragile desires. The fragility of the desires of some agents simply reflect a vigilant exercise of their capacity to adjust their desires in light of experience, information, and reflection, in an environment in which they are constantly being challenged to question whether or not their desires really do make sense. Think, for example, of the student who goes off to college and then comes home every couple of months with a completely new set of enthusiasms and ideals. Perhaps when she comes home on one occasion she has become a vegetarian and a socialist, the next time she has become a vegan and a communist, the next time she is eating shellfish and she is a liberal, and then the next time she's back to eating meat and she has become a libertarian. In other cases, however, the fragility of the desires an agent has may simply reflect the passage of time. Perhaps a certain agent's desire to watch morning television will simply disappear if for some reason she doesn't watch morning television for a few days.

With this distinction between robust and fragile desires in place, we can now offer what seems to us to be a similarly clear and intuitive suggestion about what it might mean for a desire to be an agent's own, and one with which she can be identified. Our suggestion, again unsurprisingly, is that a desire is an agent's own, or one with which she can properly be identified, if it is robust. An agent identifies with a desire just in case she sees that this is so. A desire that is fragile, by contrast, is one which is not an agent's own, and nor is it one with which she can properly be

identified or identify herself with. The appeal of this suggestion too is, we hope, plain, but, in case it isn't we offer the following.

The robust desires that agents possess help constitute their distinctive personalities. The point is perhaps most obvious in the case of the desires that agents have about how they are to live their own lives from day to day. For example, if the desires of those with whom we choose to live our lives were not robust—if they didn't have relatively stable desires to watch certain sorts of movies, read certain kinds of books, eat certain kinds of food, decorate the house in certain sorts of ways, talk to certain sorts of people, talk with those people about certain sorts of things, and so on—then it seems to us that we would not have the reasons that we have for choosing to live our lives with them in the first place. The same goes, only *vice versa*, for the desires of those we avoid because we can't stand being around them. The crucial point is that the robust desires that people have, precisely because they are robust, help to give them the personalities they have, personalities that attract us to them or repel us. And this is true whether or not those robust desires are ones that they themselves wish to be effective in action. The fragile desires that agents have about how to live their lives from day to day, by contrast, precisely because they are fragile, don't help constitute a distinctive personality. Such desires are not distinctively the agents' own, and agents cannot be identified with such desires. (Although, of course, the very fragility of an agent's desires might be a robust feature of her character with which she and others properly identify her.)

The point just made is not, however, restricted to the desires that agents have about how they are to live their own lives from day to day. When people decide which political candidates to vote for, or which charities to give to, or which people to employ, they typically make those decisions by first finding out something about the reputations that various politicians, and those who run charities, and those who apply for jobs have, for acting on certain sorts of desires rather than others. We want the politicians we vote for to have stable desires to implement the sorts of policies that we support, we want those who run the charities to which we donate money to have stable desires to distribute the money we give them in ways that we approve of, and we want those who we employ to have stable desires to do well the sorts of things that we employ them to do. Their robust desires make them the politicians, the people who run charities, and the employees, that they are, the ones who have a reputation for being orientated or disposed in certain distinctive ways. Fragile desires simply do not give us the kind of reason we require for voting for one politician over another, or for giving to one charity rather than another, or for employing one person rather than another. In a very ordinary sense, we don't know who such people are because they lack an identity.

To sum up, our suggestion in this section has run in parallel to the suggestion in the previous section. Robust desires play a crucial role in helping constitute the distinctive personalities or orientations that agents have. Because robust desires play this crucial role, it seems to us that they can be said to be an agent's own in a clear and intuitive sense. These, accordingly, are the desires with which we can identify such agents and they will be the desires with which the agents rightly identify themselves. Fragile desires, precisely because they do not so readily help constitute anything distinctive by way of personality or orientation, are thus not an agent's own in this clear and intuitive sense.

3. Descriptive versus normative questions

We said at the beginning that we thought it important to distinguish two questions: ‘Which desires (or beliefs) is an agent to be identified with, as a matter of fact?’ and ‘Which desires (or beliefs) should an agent be identified with, if she is to be free and autonomous?’ It should now be clear how the account we have given answers the first question. An agent’s robust desires and beliefs are the ones with which she is to be identified. They constitute the distinctive character of our personalities and figure overwhelmingly both in our understandings of ourselves and of others. On this account, a drug addict’s perhaps strong but ineffectual desire to be rid of her addiction, if it is a stable desire, constitutes an important aspect of who she is. But so too does addiction, and not the less because she wishes to be rid of it.

Remember we said above that robust desires might be robust for many different reasons. At one extreme, the robust desires that certain agents have are robust because the agents have and exercise the distinctive capacity that they have as rational creatures to adjust their desires in light of experience, information, and reflection. At the other extreme, however, the robust desires that certain agents have are robust precisely because they either lack or fail to use the capacity to adjust their desires rationally in the face of experience, etc., because (say) they are the product of drug addiction, or childhood trauma, or some other psychological malady. In these cases, the robustness of the desires reflects the agent’s insensitivity to the reasons she has. That rational and irrational desires alike might be robust reflects the fact that robustness, as such, is not a feature of desires that carries normative significance all by itself. The mere fact that certain desires are an agent’s own, or that the agent can be identified with certain desires, in the sense we have defined, is neither here nor there from the normative point of view.

Indeed, as we also said above, there seems to be no general requirement of rationality that agents have robust desires. Fragile desires are not, as such, defective. Agents who find themselves in an environment in which they are constantly being challenged to question what they want may well find that their desires rightly fragile as a result of their vigilant exercise of their capacity to adjust their desires in light of experience, information, and reflection. Agents are, presumably, required to exercise such capacity as they have to acquire and sustain rational desires, but this will augur in favor of certain desires being their own—that is to say, certain desires being robust—only if they happen to find themselves in circumstances in which the consistent exercise this capacity has stable desires as an upshot. Those who find themselves in situations that relentlessly provide them with good reason to change their desires will, if they are rational, find themselves with a diminished set of stable desires.

Unsurprisingly, much the same can be said about which beliefs are an agent’s own. As we said, robust beliefs might be robust for many different reasons. At one extreme, the robust beliefs that certain agents have are robust because they are so evidentially well supported by everything else that those agents believe. At the other extreme the beliefs that certain agents have might be robust for completely non-rational reasons, because (say) the agents are engaged in wishful thinking, or because they are brainwashed, or because they systematically make certain sorts of errors in their evaluation of evidence. Robustness, as such, is thus a fact about an agent’s beliefs that carries no normative significance all by itself. The mere fact that certain views or beliefs are an agent’s own, or that the agent is to be identified with certain views or beliefs, in the sense we have defined, is neither here nor there from the normative point of view.

Indeed, as with robust desires, there seems to be no general requirement of rationality that agents have robust beliefs at all. Fragile beliefs are not, as such, defective. Agents who find themselves in a constantly changing evidential environment will find themselves with fragile beliefs simply in virtue of their vigilant exercise of their capacity for reasoned belief revision. Our ability to answer the question ‘Which beliefs should be an agent’s own?’ is thus radically dependent on the circumstances that particular agents face. Agents are, of course, required to exercise such capacity as they have to be sensitive to incoming evidence, but this will augur in favor of certain beliefs being their own—that is to say, their having certain robust beliefs—only if they happen to find themselves in circumstances in which, by the vigilant exercise of their capacity to be sensitive to incoming evidence, stability in the degrees of their beliefs is the upshot.

4. Valuing, freedom, and autonomy

Let’s return to the hierarchical account of what it is for a desire to be an agent’s own. We said at the beginning that, as we see things, the crucial question is not whether there is a sense of a desire being an agent’s own, or identification with desires, which is accurately captured by that account, but rather whether the work that is supposed to be done by the idea of a desire’s being an agent’s own, in the hierarchical sense, should be done in that way. We asked whether we should appeal to the idea of a desire being an agent’s own, in the sense picked out by the hierarchical account, in order to explain what valuing is, or what freedom is, or what autonomy is.

Although we cannot argue the point at any length here, we think that the lessons we have learned in spelling out the unified account of what it is for desires and beliefs to be an agent’s own suggest that that work might be better done in another way. For we have seen not just that the fact that certain desires and beliefs are an agent’s own, in the unified sense that we have spelled out, is of no normative significance in and of itself, but also that a very clear account can be given of when desires and beliefs being an agent’s own, in that sense—and indeed of when desires and beliefs not being an agent’s own, in that sense—does have normative significance.

Specifically, to the extent beliefs and desires are subject to rational evaluation, we are in a position to distinguish between the desires and beliefs that an agent is to be identified with, as a matter of fact, and the desires and beliefs that, were she rational, she would be identified with (assuming she is in circumstances that allow her to have stable beliefs and desires without being irrational). The latter beliefs and desires, unlike the former, are the ones that she should have and they are the ones she has, and acting on which, constitute her as a free agent. It is this sensitivity of the desires and beliefs on which an agent reasons and acts to the reasons she has that makes for the difference between the actions she performs freely and those she performs merely because she is in the grip of certain of her desires and beliefs.

In saying this we are supposing, with Bratman, that the distinctive kind of agency enjoyed by temporally extended beings like us requires forming intentions and acting on plans over time in ways that are possible only when one in fact has more or less stable desires and beliefs. The upshot, as we see things, is that such agency is possible only when one has desires and beliefs with which one can properly be identified in our sense. But we are also supposing that one’s identifying with certain of one’s desires (or beliefs), and having an intention or plan to treat one’s desires (or beliefs) as reason-providing, is neither required, nor enough, to make an agent with

stable beliefs and desires a free agent. An agent who has plans and intentions concerning what she is to do, whether or not she has higher-order intentions to treat certain of her desires as reason-providing, will be a free agent so long as and only so long as her intentions, and the beliefs and desires on which she acts, are in fact appropriately sensitive to the reasons she has for believing, desiring, intending, and acting as she does (whatever those reasons are). An agent can therefore be free, we think, without having the sort of higher-order attitudes towards her own desires and beliefs that are taken as essential by those who embrace a hierarchical view.

Consider an agent who does what she wants, when she wants, how she wants, without hesitation and without exercising any sort of control over herself.¹⁷ The difference between an agent of this sort, and a free agent, is not, we think, a matter of the latter having higher-order desires (or plans or intentions) concerning her desires (or beliefs) but instead a matter of her having and acting on intentions, desires, and beliefs she has reason to have. Having desires and beliefs of one's own, in our sense, is required if she is to be an agent, but that is because, absent fairly robust beliefs and desires, she will not be in a position to form the sort of intentions and plans that constitute one as an agent over time. But such desires and beliefs need not be ones towards which the agent has any higher-order attitudes at all. And if the agent does happen to have such higher-order attitudes, those will not be enough to constitute her as a free agent if either those higher-order attitudes, or the lower-level attitudes that are effective in action, are not appropriately responsive to her reasons. What is crucial to freedom, we are thinking, is being an agent who acts on the beliefs, desires, and intentions that are appropriately responsive to the reasons she has.¹⁸

If this is right, then being free does not require having a view about the value of what one does. Moreover, if one does have such a view, it does not require acting on that view. Consider the example of Huck Finn, who does have views about the value of what he is doing, but who acts contrary to those views.¹⁹ Huck believes that Jim is Miss Watson's property, and that in helping him to escape he therefore makes things worse rather than better. But in helping Jim Huck is clearly being responsive, in his desires, intentions, and actions, to the evidence available to him of the common humanity he shares with Jim. Huck, we think, acts freely, despite his failure to endorse the desires on which he acts, precisely because his actions, and the beliefs, desires and intentions on which they are based, are responsive to the reasons that he has.²⁰

At the same time, though, we think that autonomy (as contrasted with freedom) does require that an agent that both has, and acts on, her values. Autonomy requires a kind of self-conscious

¹⁷ We have in mind the "wanton" Frankfurt describes in "Freedom of the Will and the Concept of a Person,"

¹⁸ We leave out any specific account of what all is required to count as being appropriately responsive to one's reasons, but we are assuming that having a higher-order attitude is not always required. At the same time, we suppose that having such attitudes may well often play an important role in helping an agent be appropriately responsive to her reasons. Indeed, having them may well be, itself, an appropriate response to one's reasons, in cases where one has reason to have attitudes towards one's attitudes.

¹⁹ Jonathan Bennett, "The Conscience of Huckleberry Finn," *Philosophy* (49) 1974, pp.123-34

²⁰ Of course there are some nice complications here, as it might be that the things Huck misguidedly believes provide him with some reason not to act as he does. If so, then this is a case in which the reasons Huck has are conflicted in a way that they would not be if his beliefs were different. (We take no stand on this here.) Still, even if they did provide some reason, since the weight of the reasons tells decisively in favor of Huck's acting as he does, his responsiveness to this fact would still, we think, constitute him as a free agent.

control in light of one's values that need not be present in order to be acting freely. So we think that both Watson and Bratman are on to something crucial when they give pride of place to what the agent values. However, we are inclined to embrace a view of what it is for an agent to value something that is more cognitive than what Watson and Bratman have in mind. And we think that autonomy presupposes the sort of freedom we described above, so that even an agent who successfully acts on her values will be, in our sense, autonomous, only if her values—as well as the desires, beliefs, and intentions on which she acts—are responsive to the reasons available to her.

With this in mind, let's now consider what it is for an agent to value something. As we see things, an agent values something when she believes, of the object of her desire, that it is good (or desirable). On this view, valuing is not the same as merely having the belief that something is good (or desirable), since absent a desire for it, she does not count as valuing it.²¹ Nor is valuing the same as desiring, or even desiring that a desire be effective, since absent the belief that the object of the desire is valuable, she does not count as valuing it.

Just what is involved in thinking that something is good (or desirable) is, of course, wildly contentious. Taking a lead from Bratman, one might hold that an agent who treats a desire as reason-providing is, in virtue of that, thinking of the desire, or the object of the desire, or the satisfaction of the desire, as good. (Just which would depend, we suppose, on what it is to treat a desire as reason-providing.) Alternatively, one might hold that believing something is good (or desirable) is to believe that a desire for it would appropriately survive in light of experience, information, and reflection.²² Or one might advance some other account of what it is to believe that something is good (or desirable). But whatever it is to judge something good (or desirable), we assume that an agent's desires are subject to certain standards of rationality. So it seems to us very plausible to suppose that agents who value something are committed to the idea that their desire for that thing would appropriately survive in the light of experience, information, and reflection. In any case, our understanding of valuing differs crucially from the hierarchical account, according to which an agent values (roughly) whatever, and only what, is the object of a desire that enjoys some higher-order endorsement.

The reason for preferring our more cognitive account of what it is for an agent to value something is, we think, simple. To bring it out we will suppose that believing something is desirable is just a matter of believing a desire for it would survive in light of experience, information, and reflection. With that assumption in place, imagine an agent who does, in fact, have a desire that she desires to be effective. But suppose as well that she thinks that one or the other or both of these desires would not withstand the exercise of her rational capacities (i.e., it would not appropriately survive in light of experience, information, and reflection). Is it credible to suppose that such an agent values the objects of such desires? We think not, given that (by hypothesis) she sees them as unable to sustain rational scrutiny, and so does not believe their objects to be valuable. If this is right, however, then valuing is not merely a matter of desiring (or desiring to desiring, or having a policy to treat as reason-providing, if this last does not involve believing that the objects of the desire are good (or desirable)) but must rather be a

²¹ We therefore reject the purely cognitive account of valuing argued for by Michael Smith in *The Moral Problem* (Oxford: Wiley-Blackwell, 1994) Chapter 5.

²² See *The Moral Problem* for a defense of this view.

matter, at least in part, of believing. Specifically, it must be a matter of believing of certain desires that they are for what is desirable.

Once we have on board an understanding of what it is for an agent to value something, we are in a position to capture one sense of a desire (or belief) being one's own that we think animates the hierarchical account. This is the sense in which some of our desires (and beliefs) are distinctive in being such that one is willing to stand behind them; not merely to acknowledge that one does desire (or believe) certain things, but to *claim* them as one's own, which is a matter of valuing them. In these cases the desires (and beliefs) that we claim as our own are themselves the object a higher order desire to have them and a belief that having them is valuable. A special case of this is the case of valuing our valuing as we do. In these cases, one doesn't just believe of certain desires one has that their realization is valuable – that is the original valuing -- but one also desires to have that belief and those desires themselves, believing that the realization of these higher order desires is itself valuable – that is the valuing of the valuing. In the general case of claiming ones attitudes as one's own, and in the special cases where one's values are in view and claimed as one's own, the claiming as one's own is bound up both in higher order states of desiring and believing valuable. Importantly, when ones actions, as opposed to one's attitudes are what are at issue – when the question is which actions are one's own (in the normatively laden sense) higher order attitudes need not be in play, though one must both be acting on one's desire so to act (under the circumstances) and believe so acting to be valuable (as in worth doing under the circumstances).

At the same time, it seems to us that a certain natural picture emerges of what autonomy involves. First, for an agent to be autonomous, she must have the sort of beliefs that mean she counts as someone who values acting in certain ways (which involves not simply desiring to act in those ways but also seeing acting in those ways as valuable). Second, in valuing as she does, she must actually be responding to the reasons she has, and in acting as she does, she must be acting both in accord with her values and because of what she values. On this account agents that are merely desiring and not valuing cannot be acting autonomously. But neither can agents who are valuing, and even acting on their values, if either their actions are not responsive to their values or their values are not responsive to the reasons they have. Genuine autonomy comes, we are suggesting, only when one is acting according to, and because of, one's values, and one's values are themselves responsive to the reasons one has.²³

If we are right, an agent can genuinely value various actions or outcomes or objects without having any higher-order attitudes towards her own beliefs or desires. Of course, in the normal course of things such agents presumably do become conscious of their own beliefs and desires, will come to have desires and beliefs concerning them, and will also likely find themselves valuing some of those beliefs and desires and not others. As a result, as Bratman emphasizes, this consciousness will put them in a position to recognize practical problems of coordination and adjustment that prompt solutions that can play a central role in constituting an agent both as autonomous and as the particular agent she is.²⁴

²³ This is compatible, though, with their being mistaken about what is valuable, if the reasons they have are misleading. What is required is a certain kind of responsiveness to reasons, but it is a responsiveness that is compatible with getting things wrong.

²⁴ See Bratman's "Autonomy and Hierarchy."

Specifically, we think Bratman is right that agents will find themselves facing two kinds of problems, some arising from tensions among their beliefs and desires (he calls this the problem of self-management), some arising an over-supply of apparently valuable options (he calls this the problem of underdetermination).²⁵ Bratman thinks that both kinds of problems are solved by an agent adopting a higher-order intention that structures her practical reasoning by having her treat some considerations as reason-providing and others as not. But, as we see things, the solutions will not have addressed the problems in a way that preserves agents' autonomy unless, in embracing them, agents (i) are successfully responding to the reasons they have and (ii) are, in doing so, acting in light of their values.

Understood in this way, it is worth emphasizing, autonomy does not require that the agent's beliefs be true, or that the objects of desire actually be valuable, but what it does require is significantly different than what needs to be in place for an agent to be acting on beliefs or desires that she (as it happens) desires to be effective. At the same time, though, this account echoes the hierarchical insight that autonomy is tied to having the capacity to act effectively on the basis of one's own values, even as we understand what it is for a value to be one's own quite differently.

Geoffrey Sayre-McCord
Department of Philosophy
University of North Carolina, Chapel Hill

Michael Smith
Department of Philosophy
Princeton University

²⁵ "Autonomy and Hierarchy"