

DESIRES...AND BELIEFS...OF ONE'S OWN

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 be identified with certain of her desires rather than others. The aim of such work varies. Some suggest that an account of what it is for a desire to be an agent's own provides us with an account of what it is for an agent to value something. 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 be free or autonomous.

According to the most popular account of what it is for a desire to be an agent's own, an account developed at great length by Harry Frankfurt, a desire is an agent's own if she has a higher order desire that that desire be effective in action, where that higher order desire is in turn 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. Though this Frankfurtian account of what it is for a desire to be an agent's own has many supporters, we think that there are at least two problems with it, problems which inspire us to look for an alternative.

The first problem is that it 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 view, or a certain belief, is an agent's own. Agents can, in other words, be identified more or less not just with their desires, but also with the views they hold, or with their beliefs. But it is unclear how we might generalize the Frankfurtian 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 view or a belief is

an agent's own just in case it is one which the agent wants herself to have. That sounds, if anything, like wishful thinking, not at all like identification. The second problem with the Frankfurtian answer is that it does not allow us to distinguish between what seems to us to be a descriptive interpretation of the question 'What is for a desire to be an agent's own?' and a more normative interpretation of that same question. In terms of identification, the problem is that it does not allow us to distinguish between the questions 'Which desires (or beliefs) is an agent identified with as a matter of fact?' and 'Which desires (or beliefs) should an agent be identified with?'

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. Though our paper might be read as an attack on the Frankfurtian account of what it is for a desire to be an agent's own, we do not intend it to be taken in that way. For all that we say here the Frankfurtian account might well provide us with a sense in which desires are an agent's own, and hence with a concept of identification. We take no stand on this issue. In any case, 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 desires, which is accurately captured out by the Frankfurtian 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 sense picked out by the Frankfurtian account, needs to be done in that way. In other words, do we need to appeal to the idea of a desire being an agent's own, in the sense picked out by the Frankfurtian account, in order to explain what valuing is, or what freedom is, or what autonomy is? We will make some all too brief comments about this issue at the very end of our paper.

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. We begin by focussing 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 subjects have in the propositions they believe, and there is how stable the degrees of belief in the propositions subjects believe is under the impact of incoming information and reflection. We will consider these two features in turn.

The first feature we mentioned is the degree of belief a subject has in the propositions she believes. For example, a subject might think it very unlikely that the sun will explode tomorrow, but very likely that it will rain tomorrow, and even more likely that there will be a football match tomorrow. This kind of difference in a subject's 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 thus, as we might put it, 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 in the propositions subjects believe is under the impact of incoming information and reflection. For example, though a subject might have the very same degree of belief in the proposition that the Sydney Swans is the most talented Australian Rules football team and the proposition that her son is a responsible supermarket employee, her degree of belief in the proposition that her son is a responsible supermarket employee might be very stable under the impact of incoming information and reflection, whereas her degree of belief in the proposition that the Sydney Swans is the most talented Australian Rules football team is more unstable. In other words, though perhaps nothing in the way of incoming information would change her degree of belief in

the proposition that her son is a responsible supermarket employee all that much — as we might say, the evidence is already in — all sorts of incoming information would radically change her degree of belief in the proposition that the Sydney Swans is the most talented Australian Rules football team. Whereas facts about subjects' degrees of belief are, as we saw, fixed synchronically by how much they would bet on the propositions they believe under circumstances of forced choice, facts about how stable the degrees of belief in the propositions subjects believe is under the impact of incoming information and reflection are plainly fixed diachronically. It is this diachronic feature of beliefs that will be important in what follows.

Let's call beliefs that are stable under the impact of incoming information and reflection 'robust', and those that are not stable under the impact of incoming information and reflection 'fragile'. Note that beliefs might be robust or fragile for a whole range of quite different reasons.

At one extreme, a belief might be one which a subject has because she has thought about the subject matter carefully, or has done some empirical investigation, and, as a result, she can see a whole host of interconnected reasons for thinking that the proposition she believes is true. In other words, the belief is robust because it is, by the subject's lights, evidentially extremely well supported, so well supported that it stands near the core of her web of belief. Such contrary evidence as comes in would be seen as requiring adjustments not with the belief itself, but elsewhere in the subject's web of belief. At the other end of the spectrum, however, a subject's belief might be robust because she is completely irrationally disposed to cling to her belief. The belief might be the product of wishful thinking, say, so that no matter what countervailing information comes in, the subject 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. And there

will be cases in between as well. A belief might be robust not because it is the product of wishful thinking, and not because the subject has thought about it and seen so many reasons for assigning the proposition believed the degree of belief she assigns it, but rather because she isn't very good at thinking through evidence and is vulnerable to certain systematic errors, or perhaps because she is too lazy to revise her beliefs, or for some similar reason.

The same point applies to fragile beliefs. Some beliefs that subjects possess are fragile because, existing as the subjects do in a constantly changing evidential situation, the changes in their beliefs simply reflect the vigilant exercise of their capacity for reasoned belief revision. At the other extreme, some subjects' beliefs are fragile because the subjects in question simply forget what it was that they came to believe in the first place, or because the subjects are prone to make all sorts of random 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 a subject's own, or one with which she can be identified. Our suggestion, perhaps unsurprisingly, is that a belief is a subject's own, or one with which she can be identified, if it is robust. A belief that is fragile, by contrast, is one which is not a subject's own, or is one with which she cannot be identified. 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 subjects have, no matter how their robustness is explained, would seem to constitute the distinctive world-view that those subjects have. The point is

perhaps most obvious in the case of the beliefs that subjects have about their own past. For example, if the beliefs of those intimates with whom we wish to relive our most cherished moments were not relatively stable under the impact of incoming information and reflection, then we would not have the reasons we have for choosing to interact with them in the first place. The same goes, only vice versa, for the beliefs of those who we avoid because we can't stand to relive the moments we shared with them. The crucial thing about each of these group's beliefs about their past is thus that, being robust, they help to give them their distinctive perspective on events in their past. This is what it means to say that these beliefs are especially their own. The fragile beliefs that subjects have about their past, by contrast, precisely because they are fragile — think of someone who is forever reinventing their past, or subjects whose memories wax and wane — cannot, in this sense, be said to be the subjects' own because they are not beliefs that constitute the subjects' distinctive perspective on events in their past. Given that the perspective of such subjects on the very same events changes with their changes in belief over time, they lack a distinctive perspective on events in their past.

The point just made is not restricted to the beliefs that subjects have about their own past. When people decide who to get to repair their car, for example, or which cooking teacher's classes to attend, or which clergy's sermon to listen to on a Sunday morning, they typically make those decisions by finding out, *inter alia*, about the expertise that various car repairers and cooks and clergy have. If the relevant beliefs of car repairers and cooks and clergy — beliefs about how to repair cars, about how to cook, and about the doctrines of the one true religion — were not relatively stable under the impact of incoming information and reflection, then, again, it seems to us that people simply would not have the reasons that they have for getting their car repaired by the particular car repairer that they thereby choose, or for attending the cooking classes that they thereby

attend, or for listening to the particular sermons they thereby decide to listen to on a Sunday morning. Again, then, the crucial point is that the beliefs of the car repairers and the cooks and clergy in question help to give them their distinctive perspective on that part of the world with which they are especially concerned. Their robust beliefs make them the car repairers, the cooks, and the clergy, that they are, the ones who thereby acquire the distinctive reputations they have for having the particular area of expertise that they have, reputations that make us choose one as opposed to another. Car repairers or cooks or clergy with fragile beliefs about how to repair a car, or how to cook, or what the doctrines of the one true religion are, precisely because they have fragile beliefs, don't have such a distinctive perspective, and hence don't acquire such reputations. In this sense, those with such fragile beliefs lack an identity as car repairers, or cooks, or clergy.

Here is another example, this time closer to home. A distinguishing feature of certain philosophers— part of what makes them who they are, from our collective point of view — is the fact that they are prepared to stand up for the distinctive philosophical beliefs that they hold, defending them against such arguments as are put forward by those who have contrary beliefs, or those who are as yet uncommitted. If their distinctive philosophical beliefs were not by and large robust under the impact of incoming information and reflection then we wouldn't find it as profitable to talk philosophy with them in particular. Moreover, what goes for those with distinctive philosophical beliefs goes for those with distinctive beliefs about other controversial subject matters in the arts and sciences as well. The professional identity of academics, too, then, would seem to be largely constituted by the fact that certain beliefs in their area of professional expertise are robustly held.

To sum up our argument so far, robust beliefs are those beliefs subjects have that are stable under the impact of incoming information and reflection. The robust beliefs that subjects have play a crucial role in helping constitute the distinctive perspective that those subjects 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 subject's own in a clear and intuitive sense. They are beliefs with which we can identify those who possess them. Fragile beliefs, precisely because they do not help constitute anything distinctive by way of a perspective on the world, are not a subject'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 subjects have that a certain proposition be true. For example, I might desire very strongly indeed that my children fare well as they go through their lives, whereas my desire that I watch morning television on a regular basis, though strong, might not be nearly so strong. And second, there is how stable the strength of subjects' desires are under the impact of incoming information and reflection. For example, my desire that I watch morning television on a regular basis and my desire that I play guitar on a regular basis might be of much the same strength — in a forced choice situation I might find it difficult to choose — but my desire that I watch morning television on a regular basis might be fragile, in that it will wane, and perhaps even disappear altogether, over time, whereas my desire that I play guitar on a regular basis is robust, in that it will remain.

Let's call desires that are stable under the impact of incoming information and reflection 'robust', and those that are not stable under the impact of incoming information and

reflection 'fragile'. Note that desires might be robust or fragile for a whole range of quite different reasons.

At one extreme, a desire might be robust because it coheres so well with the other things that a subject desires, something of which she is in turn reflectively aware. Perhaps the desire in question doesn't just cohere with her other desires, in the sense of not conflicting with them, but is the sort of desire the possession of which helps her to make perfect sense of the fact that she desires a whole host of other things that she desires. Perhaps it is a desire for something very general, the desiring of which makes her more specific desires much more intelligible than they would be otherwise. Her desire stands near the core of her web of desire, so that if the subject were to be led, on the basis of reflection, to suppose that her desire set as a whole needed adjustments to make it more coherent then those adjustments would seem to be needed to be made elsewhere in her desire set. At the other end of the spectrum, the desire might be robust for completely non-rational reasons, because (say) it is the product of drug addiction, or childhood trauma, or some other psychological malady. Moreover, 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 subject has reflected and seen that the desire coheres so well with the other things that she desires, but rather because she isn't very good at thinking about which desires fit coherently together and is vulnerable to certain systematic errors, or perhaps because the desire is itself an entrenched habit, or for some similar reason.

The same point applies to fragile desires. The fragility of the desires of some subjects simply reflect their vigilant exercise of their capacity to integrate their desires and aversions into a coherent whole in an environment in which they are constantly being challenged to question whether or not their desires really do form a coherent whole.

Think, for example, of the freshman college student who comes home every couple of

months or so with a completely new set of ideals, ideals formed after exposure to the new ideas and role models to which they have been exposed at college. Perhaps when they come home on one occasion they have become a vegetarian and a socialist, and the next time they have become a vegan and a communist, and the next time they have decided that eating shellfish is okay and they have become a liberal, and then the next time they're back to eating meat and they've become a libertarian. In other cases, however, the fragility of the desires subjects have may simply reflect the passage of time. Perhaps a certain subject's desire to watch morning television will simply disappear over time if she doesn't watch morning television for a while.

With this distinction between robust and fragile desires in place, we can now offer what seems to us to be a clear and intuitive suggestion about what it might mean for a desire to be a subject's own, or one with which she can be identified. Our suggestion, again unsurprisingly, is that a desire is a subject's own, or one with which she can be identified, if it is robust. A desire that is fragile, by contrast, is one which is not a subject's own, or is one with which she cannot be identified. The appeal of this suggestion too is, we hope, plain, but, in case it isn't we offer the following.

The robust desires that subjects possess help constitute their distinctive personalities. The point is perhaps most obvious in the case of the desires that subjects have about how they are to live their own lives from day to day. For example, if the desires of those intimates with whom we choose to live our lives were not stable under the impact of incoming information and reflection — 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 people, and talk 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 thus that, precisely because they are robust, the robust desires that people have help to give them the distinctive personalities they have, personalities that attract us to them or repel us. The fragile desires that subjects have about how to live their lives from day to day, by contrast, precisely because they are fragile — think of those who are forever changing their mind about what to eat, how to arrange the furniture, what to read, who to talk to, what to talk about, and so on — don't help constitute a distinctive personality for their subjects. Such desires cannot be said to be distinctively the subjects' own. Subjects cannot be identified with such desires. Indeed, in a very obvious sense, such subjects lack an identity.

The point just made is not, however, restricted to the desires that subjects 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. After all, we want the politicians we vote for to have stable desires to implement the sorts of policies that we want implemented, and 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. Again, then, the crucial point is that the robust desires in question of politicians and those who run charities and those who seek employment help constitute their distinctive personalities. 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 to implement

certain policies, or to spend donated money in certain ways, or to spend one's time at work engaged in certain sorts of activity, precisely because they are fragile, simply don't suffice to constitute the distinctive kind of orientation or disposition that would give us the kind of reason we have 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 this sense, it seems, those with fragile desires can be said to lack an identity as politicians, or people who run charities, or employees.

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

3. Descriptive versus normative questions

We said at the beginning that we thought that there was a descriptive and a normative interpretation of the question 'Which desires are a subject's own?' It should now be clear how the account we have given of what it is for subjects' desires to be their own allows us to make this distinction.

Remember we said above that robust desires might be robust for many different reasons. At one extreme, the robust desires that certain subjects have are robust because they cohere so well with the rest of their desires. Robustness in this case therefore reflects the fact that the subjects have and exercise the distinctive capacity that they have as rational

creatures to have psychological states that cohere with each other. At the other extreme, however, the robust desires that certain subjects have are robust for completely non-rational reasons, because (say) they are the product of drug addiction, or childhood trauma, or some other psychological malady. Robustness, as such, is thus not a fact about a subject's desires that carries any normative significance all by itself. As a consequence, the mere fact that certain desires are a subject's own, or that the subject can be identified with certain desires, in the sense we have defined, is neither here nor there either from the normative point of view.

Indeed, as we also said above, there would seem to be no general requirement of rationality that subjects have robust desires: in other words, fragile desires are not, as such, defective from the rational point of view. For subjects who find themselves in an environment in which they are constantly being challenged to question whether or not their desires really do form a coherent whole — remember the freshman college student — may well find that their desires are rightly fragile as a result of their vigilant exercise of their capacity to integrate their desires and aversions into a coherent whole. As a consequence, when we ask the question 'Which desires should be a subject's own?' then whether or not we are able to come up with any answer at all would seem to be radically dependent on the circumstances that particular subjects face. Subjects are, of course, required to exercise such capacity as they have to acquire coherent psychological states, but this will augur in favor of certain desires being their own — that is to say, their having certain robust desires — only if they happen to find themselves in circumstances in which, by the vigilant exercise of their capacity to have coherent psychological states, stability in the strength of their desires is the upshot.

Unsurprisingly, much the same can be said about which beliefs are a subject's own. As we said above, robust beliefs might be robust for many different reasons. At one

extreme, the robust beliefs that certain subjects have are robust because they are so evidentially well supported by everything else that the subject believes. At the other extreme the robust beliefs that certain subjects have are robust for completely non-rational reasons, because (say) the subjects 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 not a fact about a subject's beliefs that carries any normative significance all by itself. As a consequence, the mere fact that certain views or beliefs are a subject's own, or that the subject can 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 we also said above, there would seem to be no general requirement of rationality that subjects have robust beliefs at all: in other words, fragile beliefs are not, as such, defective from the rational point of view. For subjects who find themselves in a constantly changing evidential environment will find themselves possessed of fragile beliefs simply in virtue of the vigilant exercise of their capacity for reasoned belief revision. As a consequence, when we ask the question 'Which beliefs should be a subject's own?' then whether or not we are able to come up with any answer at all would seem to be radically dependent on the circumstances that particular subjects face.

Subjects 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. Unfinished business

Let's return to the Frankfurtian account of what it is for a desire to be a subject'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 out 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 sense picked out by the Frankfurtian account, needs to be done in that way. In other words, do we need to appeal to the idea of a desire being an agent's own, in the sense picked out by the Frankfurtian account, in order to explain what valuing is, or what freedom is, or what autonomy is?

Though 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 a subject'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 a subject'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, we have seen that both beliefs and desires are subject to certain rational requirements, and that it is these rational requirements that underwrite our normative perspective on whether or not certain desires or beliefs should or should not be an agent's own.

With this lesson in mind, let's now consider what it is for a subject to value something. Given that we have seen that a subject's desires are themselves subject to certain rational requirements, notably, the requirement of coherence, and given that we have also seen that which desires subjects will be disposed to acquire through the vigilant exercise of their capacity to acquire desires that accord with the requirement of coherence will depend on their informational environment — think again of the way in which the

freshman college student acquires new desires and loses old after exposure to new ideas and role models — so it seems to us plausible to suppose that for subjects to value something is, very roughly, for them to believe that that is what they would want if they had a maximally informed and coherent desire set. The reason for preferring this account of what it is for an agent to value something to the Frankfurtian account is then, we think, simply stated. For suppose that a subject were to believe that she would not have the desires that are supposed to be their own, according to the Frankfurtian account of what it is for desires to be an agent's own, if they had a maximally informed and coherent desire set. Is it yet credible to suppose that such a subject could value the objects of such desires? That is extremely hard to believe, given that she would have available the resources to rationally criticize her possession of such desires. If this is right, however, then valuing would seem to be a matter of believing — specifically, a matter of believing that that certain desires would be possessed if the subject had a maximally informed and coherent desire set — not a matter of Frankfurtian desiring.

Moreover, once this is granted it seems to us that a certain natural picture emerges of what it is for a subject to be autonomous, or free. Specifically, it seems to us that subjects who are autonomous, or free, are then naturally supposed to be those who have the capacity to recognize and respond rationally to such values as there are. In other words, subjects who are autonomous, or free, are those have the following two capacities: first, they have the capacity to have values that are evidentially well supported — that is, they have the capacity to form evidentially well supported beliefs about what they would want themselves to do if they had a maximally informed and coherent desire set — and, second, they have the capacity to acquire desires that accord with their values — that is, they have the capacity to acquire desires that accord with their beliefs about what they would want themselves to do if they had a maximally informed and coherent desire set.

The upshot is thus that, once we come to appreciate the fact that desires are themselves subject to certain norms of rationality, we will inevitably be led to reject the both claim that the Frankfurtian account of what it is for a desire to be an agent's own provides us with an account of what it is for an agent to value something, and the claim that that account is required to provide us with an account of what it is for agents to be autonomous and free. The Frankfurtian account might still provide us with a sense in which desires are an agent's own, of course. But if the fact that certain desires are an agent's own in that Frankfurtian sense is of any philosophical interest, then it will need to be demonstrated what that philosophical interest is.

Geoffrey Sayre-McCord

Department of Philosophy

University of North Carolina, Chapel Hill

Michael Smith

Philosophy Program

Research School of Social Sciences

Australian National University