{ 12 }

Two Senses of "Why"

TRAITS AND REASONS IN THE EXPLANATION OF ACTION

Iskra Fileva

1 Introduction

In Dostoyevsky's "The Idiot," the protagonist Prince Myshkin makes a character named Keller best man at his wedding despite the fact that Keller has written a damaging and libelous article against Myshkin. Myshkin is also described as looking for a reason to excuse another man, Lebedeff, for helping Keller with the damning article. Suppose we asked why Prince Myshkin acts the way he does, why he accepts with open arms the people who try to hurt him. The answer will be that he is very forgiving—so forgiving, indeed, that his virtue provokes enmity on the part of those of a less noble disposition. Thus, a character named Hyppolyte says he hates the Prince with his "jesuitical soul . . . of sickly sweetness." Were Myshkin of a vengeful nature, his behavior toward ill-wishers would be puzzling, and we would suspect ulterior motives behind the seemingly saintly façade. But given his character, we understand his actions.

As with Myshkin, so with real people: we often understand people's actions by seeing those actions as issuing from character, and we frequently appeal to character in explaining actions: "Why is your mom calling so often?" "She is very controlling." "Why is he still smoking despite his New Year's resolution to stop?" "He is too weak-willed." "Why is she drinking coffee at 10 p.m.?" "She is a workaholic and tends to stay up till 2 in the morning." The question I wish to discuss in this chapter concerns the precise way in which traits function in such explanations.[1] Just what do we mean when we say, "She did thus and so *because she is controlling, or a workaholic, etc.*?"

One may suppose that the "because" is purely causal and that traits are (or, at least, are underwritten by) some sort of blind causal mechanism. This suggestion is not without appeal. Thus, we sometimes say such things as, "Try as he may, he will never be able to quit smoking—he's too weak-willed" or, "I don't believe you could keep my secret even if you tried—you just don't know how to be discreet," implying that people's traits caused them to act in particular ways. Despite the seeming plausibility of this type of view, it is rather unlikely that traits are blind causal mechanisms. There are two reasons for this. First, ordinary observation suggests that people can act out of character. An ordinarily honest person may lie upon occasion, while a typically rude person may be unexpectedly kind to a stranger. This is not the main problem with the causal view, however, since traits can be causal mechanisms without being deterministic (after all, many causal mechanisms are not deterministic[2]).

There is a second and more serious difficulty: there is an important connection between traits of character and reasons for action (hence, my title). The actions we explain by an appeal to an agent's traits are, in general, motivated by reasons: if an action is not motivated by reasons, it cannot typically be explained by citing traits. For instance, we cannot say about the sleep-walker rummaging through his spouse's files that he does so "because he is curious" or about the person who has an obsessive compulsive disorder that "she washes her hands twenty times a day because she is overly cleanly." While a curious person may rummage through his spouse's files, and an overly cleanly person may wash her hands twenty times a day, the cleanly person and the curious person, respectively, will have reasons for action—not necessarily good reasons, but reasons nonetheless; by contrast, the compulsive hand-washer and the sleep-walker are impelled to act by non-rational forces, and the arational nature of their motivation makes trait-based explanations inapplicable to their cases.[3] But note that it is such things as compulsions that seem to be blind causal mechanisms, and traits appear to differ from compulsions precisely in not being such mechanisms.

The connection between traits and reasons is really twofold. For, of course, traits do not simply presuppose motivation by reasons in general; they presuppose motivation by particular sets of reasons: thus, a person who gives a birthday card to his supervisor may do so because he is kind, or he may do so because he is obsequious. Whether the former or the latter explanation applies depends on the nature of the reasons that motivate him.

But if both traits and reasons explanations indicate something about the agent's motivation, the two are very similar. Are they equivalent? Is the difference between them merely one of *façon de parler*, or do the two tell us different things? This is one of the questions I will attempt to answer in this chapter. More generally, my goal is to provide an account of the explanatory role of traits and traits' connection to reasons. I will do this in section 3. In section 4, I will extend the account by asking what traits actually are and will briefly consider the question whether we are responsible for our traits. I begin with a brief overview of the available theories of traits and their shortcomings in section 2.

2 The Theories on Offer and Why They Fail

2.1 THE BEHAVIORAL-DISPOSITIONAL VIEW

Since we have never directly seen a trait, it is easy to picture traits as homunculi of a sort—a little friendly man in the chest who pushes the controlling person to behave in a controlling manner and a little hard-working man who propels the workaholic to keep working. If you are wary of the danger of turning traits into homunculi, the most straightforward route you could take to avoiding the danger would be to eliminate the "internal" component of traits entirely and construe traits as purely behavioral dispositions. This is what Gilbert Ryle, for instance, proposes to do. On Ryle's reckoning, to say that *X* possesses some character trait is to proclaim as true a set of conditionals concerning *X*'s behavior. For example, to say that *X* is vain is to assert something like, "If it occurred to *X* that *A* would secure the admiration and envy of others, he would do *A*."[4] That is all—it is the whole explanation. If we then asked, "But why would *A* attempt to secure the admiration and envy of others in case he was placed in circumstances in which he could do so?" the only answer we would get would be: that's what vain people do.[5]

There is a serious problem with this picture: it allows actions motivated in the wrong way to count as trait-based. Multiple reasons for action will often be compatible with the type of behavior relevant to a particular trait, but only some of these reasons will be such as can properly underwrite that trait. Thus, the person from Ryle's example who tries to gain the admiration and envy of others may be vain but he need not be: he may, by contrast, suffer from an inferiority complex, and his attempts to secure admiration and envy may be due to a debilitating sense of worthlessness, rather than to an inflated sense of self-importance.[6]

Indeed, seemingly trait-like behavior may issue from forces completely outside the person's psychology. For instance, a person may consistently act so as to secure the admiration and envy of others but only because his brain is secretly manipulated by a powerful neuroscientist. If so, his behavior will not be due to vanity. Note that this problem cannot be solved in an *ad hoc* way, by offering a list of "external factors"—hypnosis, direct brain stimulation, and so on—said to be incompatible with trait-based behavior. We need an explanation of what it is about such "external" causes that makes traits explanations inapplicable.

There is another difficulty with a purely behavioral-dispositional view: while there are traits, such as nervousness or naiveté, which may be entirely unmotivated,[7] other traits, by contrast, may be applicable to an individual on purely motivational grounds and in the absence of any behavioral manifestations. Thus, a person may be secretly racist or secretly jealous. The secret racist may never say or do anything racist. If we follow Ryle and accept the thesis that traits are nothing but behavioral dispositions, we would have to deny the very possibility of secret racism or secret jealousy. But this is rather implausible. Indeed, it is likely that, in contemporary Western societies at least, most of racism is secret (possibly jealousy as well).

2.2 DESIRE-BASED VIEWS

The behavioral-dispositional view has a mirror image: a purely motivational view. It has been proposed that character consists of motivational states, such as desires and aversions. Since desire-based views link traits to motivation, such views could potentially offer resources to explain why certain actions with the wrong kind of motivation may appear to issue from a trait but, as in the vain person case discussed above, manifest a completely different trait or no trait at all. And since desires and aversions are motivational states, such accounts may in addition help explain why it could be appropriate to ascribe a trait solely on the basis of motivation. Can a desire-based view succeed?

The first thing to note is that on any remotely plausible version of the desire-based view, not just any desire could be part of a person's character; for instance, a desire for water when one is thirsty is not a desire constitutive of character. So a proponent of a desire-based view owes us an additional criterion, one that specifies which desires are part of character.

Different answers to this question could be given. One option is to take endurance to be that additional criterion. Richard Brandt, for instance, takes just this route. After criticizing Ryle's theory of traits on grounds somewhat similar to mine here, Brandt proceeds to argue that traits are enduring desires and aversions (which, together with instrumental beliefs, lead to action) (Brandt (1970, 1992).

Brandt's view presents a clear advancement over Ryle's: it affords us resources to rule out many "wrong motivation" cases, and it can also help make sense of purely motivation-based traits such as secret racism. Nonetheless, the account fails. It fails because it allows pathological behavior to qualify as trait-based: a strong and persistent aversion to heights or an intrusive desire to wash one's hands frequently will usually be symptoms of acrophobia or of an obsessive-compulsive disorder, respectively. If so, they are not part of one's character or personality. But the persisting desires theory of traits does not allow us to explain why they are not. Once again, an *ad hoc* solution will not do: we cannot solve the problem by simply stipulating that the desires in question must, in addition to being persistent, be non-pathological. We need an account that tells us what about pathological desires disqualifies them.[8] The answer, I will argue here, has to do with the fact that an obsessive-compulsive action is not motivated by reasons and so cannot be explained by an appeal to an agent's traits.

There is another problem with Brandt's account: one and the same enduring desire can underwrite different traits. For instance, a desire for dominance might make someone aggressive, or passive-aggressive, or manipulative (there are, after all, different tactics one can use to gain advantage). Apparently, none of these traits is simply reducible to the desire for dominance, since the three are clearly not one and the same trait. This suggests that at least some traits have an ineliminable behavioral component and cannot be ascribed to a person simply on the basis of motivation, irrespective of the actions which the motivation leads to.[9]

It could be, however, that the problem here is with Brandt's particular version of the desire theory, not with desire-based theories of character in general. Other versions of the desire-based view are possible. For instance, one could take identification, rather than persistence, to be the distinctive mark of the desires that enter the constitution of a person's character. Thus, Bernard Williams tells us that for a person to possess character is for him to have "projects and categorical desires with which that person is identified" (1982, 14).

Now, there are different things one can mean by "identification" here. In the first instance, there is the ordinary idea of "identification" on which to identify means, roughly, "to recognize yourself in . . ." In this sense one may, when sober, not identify with a desire to break vases, which she gets when inebriated.[10] This view faces clear challenges: we may be deceived with regard to the desires we have and it may nonetheless be the case that those desires are an integral part of our characters. Thus, the powermonger who does not realize that her primary reason to run for office has to do with the fact she enjoys power over others is, despite all that, a powermonger, just one lacking self-insight.

There is a second idea, more popular among philosophers, on which identification is to be construed as a kind of endorsement. A strong version of this view will require active endorsement: reflecting on a desire and accepting it as your own. This version will not do: while some of the most central aspects of our characters clearly have to do with motivation we have actively endorsed in this sense (e.g., Martin Luther's "Here I stand"), other aspects, such as powermongering I am not aware of, or vengefulness I know I possess but do not endorse, may be part of our character despite lack of active endorsement on our part.

But a weaker and more plausible version of the view is possible, and a sketch of such a version can be found in Williams's (1993) writings on the related topic of moral incapacity.[11] On this weaker reading of the endorsement view, it is not necessary that I reflect on a desire and approve of it in order for it to count as part of my character; it suffices that I do not actively oppose the desire in question. If I do oppose it, it will cease to be a part of my character. Thus, suppose I have a desire to be loyal, but I set out to overcome it, say because betraying my friend would bring me a great benefit. If I make up my mind to overcome my commitment to my friend, I may still find myself unable to do so (in this sense, my desire to be loyal may endure, despite an opposing second-order attitude), but my incapacity to betray would have the status of an inner inhibition, and it will not be, properly speaking, a part of my character, or, as Williams puts it, it would be a purely psychological incapacity, not a moral one. Surely, a person who is actively trying to overcome her loyalty to a friend is no longer a loyal friend, she is no longer "identified" with her loyalty: it is not only that she does not see herself as being loyal, since the powermonger may not see herself as such either; we no longer believe she is loyal.

In at least one respect, such a proposal would be preferable to Brandt's: the former, unlike the latter, rules out many, perhaps most, pathological desires

since people do not typically identify with pathological desires—either in the ordinary or in the endorsement sense—and they often actively attempt to free themselves of such desires, perhaps even seeking help from psychotherapists to effect the changes in question. But in the end, this conception does not work either. On the one hand, the "identification" criterion is too restrictive. Not all traits are underwritten by desires one identifies with, even in the minimal sense I spelled out above. Consider again the powermonger who disapproves of her desire for power and actively seeks to rid herself of it. Before she succeeds, we may still characterize her, quite appropriately, as a powermonger, albeit one who is trying to change (she may also fail to recognize she has such a desire and act with the rationalizing belief that the reason she is seeking a position of power is that she knows she is capable of taking responsibility and of doing good).[12]

The identification criterion is, in another sense, too broad: it is possible for an agent to identify with a pathological desire.[13] In reality, this may not happen frequently, but it can and does happen. The compulsive hand-washer from the earlier example may see the desire to wash her hands a few times an hour as "her own," both in the sense that she recognizes it as expressive of who she is and in the sense that she endorses it.[14] She may hold the rationalizing belief that she acts as she does because she is overly cleanly, and she may be quite proud of this self-ascribed trait.[15] Yet, she is not cleanly, she has a disorder, and the character trait "cleanly" is inapplicable to the actions resulting from that disorder.[16] I conclude that the "identification" criterion fails as well. As we will see shortly, the problem with this version of the desire-based theory lies to a large extent in the absence of a connection between traits and reasons.

2.3 A NORMATIVE REASONS VIEW

Sophia Moreau criticizes Williams's desire-based view and proposes an alternative on which traits are construed as dispositions to respond to normative reasons, reasons that exist independently of and are not rooted in an agent's desires (Moreau 2005, 274).

I discuss in detail Moreau's objections to Williams as well as Moreau's own proposal elsewhere.[17] For present purposes, Moreau's argument can be summarized as follows: in construing character as a set of desires, Williams fails to accord normative reasons their proper role in the constitution of character. When the loyal person says, "I cannot betray my friend," she is not reporting her own desires. She is not saying, for instance, that she has some sort of psychological aversion to disloyalty. Rather, the loyal person responds to normative reasons that she perceives as independent of her desires.[18] Moreau goes on to fault Williams for collapsing the distinction between character traits and psychological conditions such as phobias: it is psychological conditions that consist of psychological, motivational facts. But character traits cannot be construed as

sets of motivational facts—they are dispositions to respond to reasons existing independently of our motivation.[19]

There is some truth to Moreau's proposal. In particular, it is true that when we act in character, we, quite often, perceive our decisions as constrained solely by normative facts independent of us and of our characters. The problem is that this perception—whenever we have it—is not the whole story: it leaves out the fact that the conclusions of our deliberations and the way in which we view the situations we find ourselves in depend on our desires and on our characters generally. While it is quite true that the loyal person who says to herself, "I cannot betray him," typically believes her inability to be grounded in normative facts independent of her, we, as interpreters of her action, need an explanation of why she believes this. Why doesn't she say to herself, "Well, I can betray him if it comes to it" instead? She may not need an answer to this latter question, but we do. After all, were she to have a different character, she may well have reasoned differently. A disloyal person, for instance, is unlikely to say to herself, "I can't do this. That would be disloyal." A reasons explanation of an action typically suffices for the agent performing the action (you have reasons for what you do that you find compelling, and that is all there is to it for you), but it does not suffice for an interpreter. I will argue here it is precisely this insufficiency of reasons explanations which constitutes the main reason we appeal to traits.[20]

Note that in rejecting Moreau's account of traits in terms of dispositions to respond to normative reasons, I do not wish to deny that there could be normative reasons that exist independently of our desires. Rather, I want to insist that character traits are best understood as the subjective conditions that predispose an agent to act on a reason she perceives, whatever the ontological status of that reason.[21]

What, then, are traits and how do they explain actions? So far, I have argued that traits are neither behavioral dispositions, nor persisting desires, nor desires one identifies with, nor, finally, are they the sort of dispositions to respond to normative reasons that Moreau argues they are. However, as we shall see in what follows, there is some truth to all of the accounts I rejected: traits do have a behavioral component, a desire-like component, and a connection to reasons, in addition to one fourth element that we shall come to. Before I turn to the nature of traits, however, I wish to discuss traits' explanatory role.

3 How Traits Explain Action

3.1 CONSISTENCY AND ITS EXPLANATORY POWER

I mentioned earlier that the behavioral-dispositional view has at least one virtue: it avoids turning traits into homunculi. I wish to point out now that this is not the only virtue of the view. Another strength of the account is that it aims to do justice to the most conspicuous feature of traits: consistency across contexts.

In order to appreciate the importance of this feature of traits, suppose that your coworker Jill is cold and unwelcoming with you specifically but not with people in general. Then unfriendliness will not be one of her traits, and it will not explain why she seems ill-disposed toward you. We call someone "unfriendly" only if we observe her behave in an unfriendly manner toward multiple people.[22] The consistency need not be perfect—even the friendliest person is bound to act disagreeably sometimes—but a friendly person will tend to be amiable most of the time, and no one can be properly called "unfriendly" on account of being surly and disagreeable with a single other person. This aspect of traits is one that the behavioral-dispositional view accommodates and makes central.

However, consistency cannot be the whole story. It cannot be, because traits tell us something not captured by consistency alone. Above, I gave arguments to the effect that traits have a motivational component: the arguments had to do with the intuition that a trait ascription requires the right kind of motivation. Here, I want to cement my case by approaching the issue from a different angle, that of action explanation, rather than trait attribution. Assume for the sake of argument that traits are just behavioral consistencies. Then traits explanations would be interchangeable with explanations that appeal to consistency. Now, suppose you ask a coworker why Jill always frowns at you. In response, the coworker says, "She frowns at you because she is inhospitable with Paul, and Sarah, and Margaret, and bank accountants, and coffee shop baristas." Clearly, this explanation will not do. While Jill can be properly described as "unfriendly" only if she behaves in a hostile manner in a variety of contexts, it would be false to say that Jill is unfriendly on one occasion because she is dour on other occasions. Here, one is inclined to reply: well, it is not consistency *per se* which has explanatory power; rather, it is consistency because of an underlying trait. But the problem is just what this means.

Before I take this question, I must note that in saying that consistency by itself is not sufficiently explanatory, I do not mean to imply that consistency has no explanatory power. The point is, rather, that consistency alone does not suffice to answer certain "why" questions regarding action, questions that we frequently answer by invoking a trait. Returning to the example above, consistency does not tell us why Jill frowns at you. Yet consistency is already explanatory, because it can help answer another question regarding action, a question such as, "Is there something about this particular case which makes the agent act as she does, or is it that the agent's present behavior conforms to a pattern?"[23] Indeed, the response to a question of this sort may be all we need to know on a particular occasion, and it may well be all you need in the case of Jill: if you learn that Jill's hostility is due to something about her, not you, you will probably come to consider the matter closed. The causes behind Jill's attitude can remain a black box: whatever their nature, they have nothing to do with you. But if our goal is to provide a theoretical account of traits, we need to go further than that. We need to shed light on the power of traits to help answer additional "why" questions.

3.2 TWO SENSES OF "WHY"

The first thing I wish to do in an attempt to untangle the issue of the explanatory role of traits is to distinguish between two senses of "why" when it comes to questions concerning action. When we ask why someone does what she does, we may mean to inquire about her reasons. This is the first sense of "why." But we may, second, mean to ask why an agent acts on one as opposed to another available reason. Traits as well as reasons can be cited in answering the first "why" question. Since the question asks for reasons, the most direct way to answer it would be to cite reasons. Yet character traits, because of their connection to motivation and reasoning, can help answer that question as well. Thus, suppose we ask why somebody gave a card to his supervisor, and we mean that as a question about reasons. We get the response, "because he is kind." This response will normally satisfy us: it tells us what type of reasons motivated our agent—in this case, the reply indicates that the agent was moved by concern for another person's well-being or a desire to relieve her stress, for instance—rather than by a selfish concern, such as a desire for promotion. The response may not tell us exactly what reasons motivated the agent since, as we noted in the beginning, traits are connected to kinds of reasons rather than to particular reasons (thus, the kind person may have said to himself, "I want to cheer her up" or "This won't really cheer her up, but at least, it will show I care."), but we are often not interested in such details.[24]

More important, character traits are uniquely suited to answering the second "why" question, the question about preferring one as opposed to another reason. This is why being loyal, for instance, cannot be simply a matter of acting rationally: both loyal and disloyal behavior can be rational.[25] When, by contrast, we seek to explain an action for which there is no rationally permissible alternative available, we do not appeal to traits. Consider an illustration. While walking down the street, Abe coincidentally finds a lost wallet with an ID and $200 cash in it. He takes the ID out, retrieves the owner's information, finds the owner, and promptly returns the wallet. In explaining why Abe acted as he did, we are likely to say, "He returned the wallet because he is an honest person." To say this is to appeal to a character trait. Why do we appeal to Abe's honesty here? Because several rationally permissible courses of action were open to him, and we need an answer to the question of why he chose the particular option he chose.[26] For we can surely imagine someone else in Abe's position taking the money and abandoning the wallet with the ID, or even taking the money and throwing the wallet with the identification card in the trash.[27] These alternative possibilities give rise to a question in the interpreter's mind: Why did Abe act the way he did rather than in some other way? Why wasn't he moved by considerations of self-interest and monetary gain, for instance? This is why an account such as Moreau's cannot ultimately succeed, despite Moreau's salutary attempt to connect traits to reasons: reasons by themselves cannot explain why an agent acted on those as opposed to other available reasons.

Now imagine, by contrast, a situation in which only one available option is rationally permissible. Thus, imagine Abe has noticed that his own wallet with his own ID and driver's license in it is missing. He decides to retrace his steps and look for the wallet. At one point, he spots the wallet lying on the ground, and he takes it. In this case, we would not seek to explain why he does what he does by citing a trait that he, presumably, possesses. (What trait could that possibly be?) Here, an appeal to traits is neither helpful nor, in a sense, possible because the agent has taken the only rationally permissible way of acting. Taking the only rationally permissible course of action open to one is simply a matter of acting rationally; it does not depend on specifics of one's character.[28]

We now see that traits and reasons explanations, while systematically connected, are not equivalent, and that the difference between them is not merely one of *façon de parler*: while both traits and reasons can answer the first "why" question concerning action, traits are uniquely suited to answering the second.

Personality psychologists have asked why we appeal to traits in explaining people's actions, and the answer they offer is different from the one I just suggested. They typically suppose that traits help us explain why different individuals act differently under similar sets of circumstances.

Trait theory in psychology has been seen as a theory of what individuates people (see Brody 1988; Funder 1994; Goldberg 1995). There is truth to the suggestion. It is likely that the need to individuate people in ways that are not merely physical ("she has green eyes," "he has dark hair," etc.) but psychological ("he is honest," "she is friendly," etc.) is, in fact, behind the evolution of trait vocabulary; but with regard to the question I am raising in the present chapter, namely, what role traits play in explaining action, this way of looking at things is not sufficiently illuminating. The reason why it is not is best given by reference to an insightful remark made by John Doris. Doris notes that the individuation of behavior is not necessarily entailed by the concept of a personality trait. He writes, "Suppose that everyone behaved honestly in a situation with strong pressures to deception; this looks to be some evidence for attributing a robust trait of honesty to each member of the population, despite the absence of individuation" (Doris 2002, 177n23). Doris is right—while trait terms are commonly used to individuate people, individual differences are not principally required in order for a trait term to be meaningful and to have an appropriate application. However, my account does not fall prey to Doris's objection in the way the view of personality psychologists does. It is, indeed, possible to imagine a community of beings all of whom exhibit a particular kind of trait-relevant behavior—for instance, much like the knights from the Knights and Knaves logic puzzles, everyone tells the truth. Intuitively, it seems that the trait term which serves to label the common propensity would still apply: all will be honest. Personality psychologists who suppose that the function of traits is to help us individuate among people cannot explain why this is so. Given my answer to the question about the function of traits, however,

I can: even if everyone behaved honestly, the trait ascription "honest" would still be applicable because it would remain the case that a rational person could behave dishonestly (without losing his or her rationality), and so a traits explanation such as "because they are honest" would be our answer to the question why none of the people imagined chooses to lie. By contrast, when we think of Abe's attempt to retrieve his own wallet, we do not look for a trait we can appeal to in order to explain Abe's action; we do not, because there are no reasons for him to act otherwise. Since traits explanations answer the question: "Why did the agent pursue this course of action rather than one of the available alternatives?" a traits explanation becomes inapplicable in case there are no rational alternatives.

Two more points must be mentioned here. First, earlier I argued that when only one course of action is rationally permissible and the agent takes it, we do not appeal to a trait in explaining why he does what he does. I must now add that when the agent whose action we are trying to explain takes none of the rational choices available and acts in a rationally impermissible way instead, we would not seek to explain why he does what he does by an appeal to a character trait either. Thus, imagine Abe stumbles across his own wallet with his own ID and driver's license in it but abandons it, or throws it in the trash. This would lead an interpreter observing Abe's behavior to ask: "Why did he do that?" But we do not anticipate that the answer to this question will have anything to do with traits. If Abe stumbles across his own wallet, ID, and driver's license but does not take them, we would probably think that he is a mentally ill person, unable to recognize his own picture ID.[29] On the view developed here, this is precisely what we must expect: as mentioned in the beginning, it is only reason-based actions that we explain by an appeal to traits.

Second, there are cases in which a person adopts one of several rationally permissible ways of acting, but we do not think a traits explanation of why he or she takes one rather than another course is called for. This may be because we think that another kind of explanation is better, for instance, one which cites the agent's mood. Thus, the correct explanation of why Abe returns the lost wallet may be that he has just gotten a promotion and is feeling particularly generous. Or the correct explanation may have to do with a situational factor, for instance, the ID in the wallet, judging by the photo, belongs to a beautiful woman.[30] Finally, there will be instances in which there is simply no good answer to the question why someone does the thing he does rather than something else. That would be the case when, for example, the agent in question is particularly fickle and unreliable—one day, he would return a lost wallet, the next day, he will avoid picking up the tab.[31]

The three cases just mentioned—the promotion case, the beautiful stranger case, and the fickle person case—raise a number of questions of their own: How do mood and situational factors affect behavior? What is the connection between mood, situation, and character? I cannot address all these questions here. Suffice

it to note that traits are not the only resource we appeal to in attempting to answer the second "why" question.

We now know how traits function and whether they answer the same "why" question which reasons answer or a different "why" question. But our account will be incomplete if we leave it at that. So far, we have described the explanatory role traits play in action explanation. But we must ask also how traits succeed to play that role. And in order to answer that, we must address the question what traits are.

4 What Are Traits and Are We Responsible for Them?

Earlier, I pointed out that the first key component of traits in general is behavioral: traits indicate regularities in a person's actions ("in general" is important: as mentioned in the beginning, not all traits have a behavioral component: some—like secret jealousy—may not have an outward expression at all[32]). But I noted also that traits cannot be *reduced* to this component.

The second main element of traits, as suggested by the arguments so far, is motivational. The connection between traits and reasons I have been emphasizing throughout this chapter comes via this connection to motivation. The motivational element of traits is a complex issue to which I cannot do full justice here, but the main point to note is that this component has to do, above all, with a person's values. The loyal person is loyal because she values certain things, while the disloyal is disloyal because she values other things. Values, in general, influence us by influencing the sorts of reasons we perceive as salient and by motivating us to act on certain reasons rather than others.[33] This is why traits tell us both what sorts of reasons routinely motivate someone and why the person has acted on those as opposed to other kinds of reasons.

We can now see why pathological desires are not a part of character: they are not expressive of a person's values (a cocaine addict cannot be said to "value" cocaine, for instance[34]) and do not incline a person to take one as opposed to another rational course of action—rather, they push her to take irrational actions. We have an intuitive grasp of this: if we see a person washing her hands five times in the course of an hour, we do not think she does so because she is cleanly, since, while we think that a rational person may value hygiene more or less, we do not think that it is possible to rationally value hygiene so much that you would interrupt dinner with family or a conversation with a friend on the phone in order to go and wash your hands again.

We are now in a good position to see why we are responsible for our characters. While many suspect that we are, no good answers to the question "why" are usually given. Bernard Williams, for example, says that "if one acknowledges responsibility for anything, one must acknowledge responsibility for decisions and actions which are expressions for character" (1982, 130). His thought appears to

be that character is who we are, and that we must assume responsibility for who we are, sort of the way adolescents are told to "own up to it." The claim has some rhetorical force (if only because it is often older and wiser people who tell us to own up to it), but Williams gives no satisfactory answer to the question why it is that we are responsible for our characters.[35] And the answer, I wish to suggest, is that we are responsible for all of the actions we do for reasons. Since actions that flow from our characters are done for reasons, we are accountable for those. And we are responsible for our characters indirectly—it is primarily via our responsibility for the reason-based actions that flow from our characters that we are responsible for our characters.

But why, one may ask, is it that we are responsible for the actions we do for reasons? If, as I have argued here, the values that comprise our characters incline us to look at situations in particular ways, do we really have the freedom to do something other than what our values incline us to do, and if not, are we responsible?

The answer is that we do have the freedom. We can look at a situation in more than one way and from more than one point of view.[36] This may not be an easy thing to do, but we do have the capacity to do it. If we did not have reflective capacities and ability to assess situations from multiple points of view, if we were trapped in the way of seeing the world underwritten by our characters, we would not be truly responsible for either our characters or the actions that flow from them.[37] This is the first half of the answer. The second part of the answer is this: once we see the situation differently, we are capable of mustering the motivation to act contrary to our character propensities—character does not provide irresistible motivation.

There is, then, some truth to Moreau's account mentioned earlier: it is true that our responsibility for our characters has to do with an ability to act on reasons independent of our desires. But it is not true that this ability is what constitutes character:[38] a person who is ordinarily ill-behaved but who today surpasses herself and succeeds in behaving kindly for reasons independent of her desire to be rude is acting out of character. And because her action is motivated by reasons, she is responsible for the action as well as deserving of praise (indeed, the out-of-character kind action may be more praiseworthy than that of a person who behaves kindly habitually, since the former type of action is a greater accomplishment[39]).

I have to make a final point before concluding this section. As indicated in the beginning, traits contain an extra element, in addition to motivation, behavioral dispositions, and connection to reasons: something akin to skill or a certain measure of success in acting. Take, for instance, a trait such as "ambitious." A person is only properly called "ambitious" if he achieves the goals he sets for himself: a private who dreams of becoming a general (and, perhaps, unsuccessfully attempts to get a promotion) is not, on that account, called "ambitious." Similarly, the trait ascription "humorous" implies more than motivation and behavioral dispositions. The humorous person not only desires to and routinely

attempts to but succeeds in making others laugh. This additional element is likely to be present in all so-called "intellectual virtues"—being broad-minded, unbiased, and so on—but it will be essential also to many moral traits, such as being tactful, as well as to virtues of taste, for instance, showing refined sensitivity or taking pleasure in what is subtle. In some cases, it may be appropriate to ascribe a skill-related trait in the absence of any motivation, for instance, a person may be properly called "wise" even if she does not want to have any insights and considers wisdom a source of suffering.[40] This is another reason why a desire-based view of character cannot succeed. And it adds an additional complication to the conditions of responsibility for character: to the extent that a trait has an element of luck, we are not responsible for it, although it is not the case that we are in no way responsible simply because there is some element of luck.[41]

Note that this general framework leaves many details unspecified. The precise degree of responsibility an agent has for his or her character and the actions that flow from it would depend on multiple factors, including how a trait was acquired. For instance, violence may have a genetic basis, stinginess may be due to growing up in times of economic depression,[42] while obsequiousness may develop as a result of living in a very hierarchical society. In each of the cases mentioned, we will need additional information in order to assess a person's responsibility for his or her character traits; for instance, in the case of aggression, we will need to find out how much of the variation is explained by genes, what other important factors besides genes influence aggressive drives, and whether and which of those factors are present in some particular case. The important point to keep in mind is that we can see the world in ways different from those in which our characters incline us to see it, and we can act in accord with those other ways.[43]

5 Conclusion

It is now time to pause and take stock. I began this account with the questions of how traits explain actions and how traits explanations relate to reasons explanations. I argued that traits explanations, while they indicate something about an agent's reasons (in virtue of the connection between a given trait and particular sets of reasons), tell us something else in addition: they tell us why the agent acted on those reasons rather than on other possible reasons. Thus, traits answer a second kind of "why" question concerning action. Next, I claimed that the motivational component of traits has to do with a person's values, and the values underlie patterns in an agent's reason-based, trait-relevant behavior, they make a person inclined to regularly act in some ways rather than other ways, although they do not determine what a person will do. And I claimed that we are responsible for our characters primarily because we are responsible for all reason-based actions, and actions done in character are done for reasons. Finally, I traced our responsibility for our reason-based actions to our reflective capacities.

I would like to make some final remarks before closing this discussion. I have argued that the motivational component of traits is best thought of as values that incline a person to act on one set of reasons rather than another. One may have residual questions regarding one sort of scenario: the Huck Finn case. One may agree that it is ultimately Huck's values that lead him to make the choice he makes but wonder whether the values influence him to act on one as opposed to another reason, as I claimed character does. Is Huck acting for reasons at all? My answer is that he is: he acts for reasons he does not fully realize he has. We are not transparent to ourselves, and sometimes, we act for reasons we don't know we have.[44] In such cases, our reasons are often worse than we think they are, but they could be better as well. Just as the powermonger who mistakenly believes she is acting on the motive of duty may nonetheless be manifesting the trait of powermongering, so too Huck, who mistakenly believes he is doing something wrong and unjust, may nonetheless be manifesting the virtues of humanity and justice.

Throughout this chapter, I have been discussing the way in which traits can illuminate action. "Can" here is important: ordinary traits explanations are very often partial and incomplete. Recall the point we made earlier about the explanatory power of consistency. If you are bothered by Jill's unfriendly attitude, and you find out she frowns at everyone but you don't learn why, this (virtue dormitiva) explanation will typically suffice for your purposes—Jill has nothing against you personally, she is just unfriendly to people, and that is all you want to know. You are not likely to inquire further into why Jill has this trait. If you are a psychologist analyzing Jill, you might. You would probably wish to know the full story behind Jill's lack of cordiality—did she, perhaps, have bad parents and grew up to be passive-aggressive? Or did she, rather, used to dream of being a ballet dancer but ended up working as a bank accountant and is now too frustrated by the course her life has taken to be friendly to anyone? You may also wish to know whether there are any people in Jill's life to whom she is friendly, and if so, who those people are. But if you are Jill's coworker, not her psychologist, having the consistency in Jill's behavior across situations pointed out to you will suffice. The upshot is that we often have neither the time nor the interest to acquire intimate knowledge of others. Our theoretical account of traits, then, is meant to tell us not so much how traits explanations function normally, but how they would function ideally, that is, if we were interested in both the nature of traits and in each other's characters.

Notes

1. For considerations of space, I will not discuss situationist challenges to the existence of character and will assume that character explanations are sometimes appropriate. I will note, however, that psychologists have moved beyond the simple person-situation dichotomy, and the consensus view, currently, is that both situational and personality variables play a role in affecting behavior. See, for instance, Mischel 2009, 284; Mischel, Shoda, and Denton 2002 (Mischel is the psychologist whose book *Personality and Assessment* ignited the debate in the 1960s); Fleeson 2004; Lucas and Donnellan 2009; and Funder 1994.

2. For instance, medications work via causal mechanisms, but a medication may or may not work.

3. Again, the point is not that psychiatric conditions determine behavior while traits do not, since psychiatric conditions need not determine behavior, for instance, a person with extreme social phobia may be able to get out of the house if the house is on fire, a point made also by Mele (2004, 78–88). My claim is that psychopathologies typically function as causal forces with no systematic connection to a person's reasons for action.

4. See Ryle 1949.

5. Margaret Gilbert calls this "the summary" view of character and proceeds to reject it. See Gilbert 2006.

6. There are often discrepancies between explicit and implicit self-image, frequently with disastrous consequences. See Creemers et al. 2012.

7. A naive or nervous action will, in the usual case, be motivated, but its being naive or nervous will not be motivated. For instance, one may naively sign a document that deprives her of inheritance or nervously look for a missing item in one's bag. Both signing the document and looking for the missing item will be motivated actions, but one will have no motif to perform these actions naively or nervously. If, on the other hand, a person feigns naiveté or nervousness, the naiveté or nervousness will be motivated, but her action will, for just this reason, be neither naive nor nervous—it will only appear so, if she is good at feigning.

8. Sarah Buss (2012) suggests that actions motivated by pathological desires are not actions we can be held responsible for, because pathological desires are incompatible with minimal human flourishing and thus undercut sanity and mental health. We can extend her suggestion to character by saying that pathological desires cannot be constitutive of character for the same reason. But this suggestion won't do: some kinds of pathological desires are quite compatible with minimal human flourishing, for instance, an acrophobic can flourish and can even avoid the negative impact of her condition by simply avoiding heights. On the other hand, character traits may have an impact on human flourishing much more significant than pathological desires, for instance, a person who harbors too much hatred or anger may be much less capable of flourishing than a person with a phobia.

9. See Antti Kauppinen's "Character and Blame in Hume and Beyond," this volume. The point also has resonances with Christian Miller's account of mixed traits. See his chapter "Mixed Traits," this volume.

10. I model this example on examples discussed by Nomy Arpaly in "Moral Psychology's Drinking Problem," this volume.

11. Note that in this paper, Williams does not seek to explain character—rather, he takes character as explanans and moral capacity as the explanandum. But if he is right about the nature of moral incapacity, his account may nonetheless help shed light on the nature of character.

12. There may be an asymmetry here between positive and negative traits—it may be that one can only be said to possess an evaluatively positive trait such as loyalty if, at minimum, one is not actively attempting to free oneself of that trait, but that the possession

of a negative trait requires something weaker and may be compatible with attempts to change that trait.

13. Sarah Buss (2012) makes this point as well.

14. There is evidence that even people who act on a hypnotic suggestion may think that they are acting for reasons. See Wilson 2004, 95.

15. Relatedly, psychiatrists have recently described a disorder consisting in an obsession with the "purity" of food. People with the condition typically feel virtuous for having it. See Hill 2009.

16. For a discussion of the difference between a trait and a disorder, with focus on OCD and scrupulosity, see Walter Sinnott-Armstrong and Jesse Summer, "Scrupulous Character and Mental Illness," this volume.

17. Iskra Fileva, *Ghostwriters of the Soul*, in progress.

18. Williams's view has been criticized both on the ground that it does not properly incorporate the role of deliberation and on the ground that it accords too big of a role to deliberation. Sophia Moreau (2005) raises the former objection; Craig Taylor (2001) raises the latter. Taylor argues that Huck's incapacity is moral but not a result of deliberation, rather, it is a brute fact about Huck (it is moral in virtue of having the right sort of connection to Huck's motivation).

19. She writes: "'I can't do that! That would be disloyal!' does not appear to me as a brute fact, a fact about myself that I am simply landed with. Rather, I see myself as answerable for this psychological fact: I believe I could be asked to give reasons for it and that I could be in error about these reasons" (Moreau 2005, 294).

20. Elsewhere, I have argued that the virtuous person perceives certain reasons as normative, although, objectively speaking, they are not; for instance, she may perceive it as her duty to do something that is, objectively speaking, supererogatory. See Fileva 2008.

21. The case is analogous to that of being moved by reasons to believe. Such reasons can be said to be, in some sense, independent of an individual believer, but the likelihood they would move a believer depends on various subjective conditions. Those conditions in the theoretical reasoning case have to do with factors such as intelligence, ability to reason logically, the presence of potential biases and prejudices, etc. In the practical reasoning case, they typically have to do with an agent's traits: our traits incline us to see reasons for action differently from the way in which a person with a different character placed in the same circumstances would see them. Later, I will suggest that when a person is moved by a normative reason independently of any prior desires, she is, in the usual case, acting out of character, although the action may signal that she is about to have a moral conversion.

22. Or, in any event, it is only then that we have a justification for calling someone "unfriendly." For, of course, upon occasion, we make trait attributions without justification, a tendency variously dubbed by psychologists: the Fundamental Attribution Error, The Correspondence Bias, The Over-Attribution Error, and the Attribution Effect.

23. Consistency may also undercut the motivation for asking additional "why" questions because if Jill behaves with you the way she behaves habitually, we are not surprised by what she does and feel no need of further explanation.

24. This is so because frequently, what we are concerned with is what sort of person someone is, and in order to learn about that we usually need to know what type of reasons motivated the agent, not what particular reasons she gave herself. On this point, see Theodore Hayes, Robert Hogan, and Nicholas Emler, "The Psychology of Character, Reputation, and Gossip," this volume.

25. One could, of course, try to argue that acting immorally is somehow irrational. On some interpretations, Kant held this view. Adrian Moore (2003, 23–30) for instance interprets Kant this way and objects on the ground that immoral action need not be irrational. I myself do not subscribe to this way of reading Kant, but whatever the truth about Kant's view, I take it as quite obvious that immoral action in general need not be irrational.

26. Note that sometimes, all we need is an intelligible explanation. Thus, Christine Korsgaard tells us that if we heard someone went from Indianapolis to Chicago with the sole purpose of buying a box of paper clips, we would not believe this and think that there must be more to the story. See Korsgaard (2009, 14). This is quite true, and it will often be the case (though this is not Korsgaard's point) that once we get an intelligible explanation, we do not inquire further.

27. Indeed, on certain conceptions of rationality, the second option may count as "more rational."

28. It is possible to question the distinction I just drew between acting from a trait and acting simply rationally. Sometimes when we speak of "rationality," we have in mind a universal characteristic of all (mentally healthy) humans. But at other times, we speak of "rationality" as though it is a character trait of individual people. We say things such as, "Warren Buffet is more rational than most of us, and this is why he is capable of buying when everyone else is selling or selling when everyone else is buying." Attributions of this kind suggest that we believe there is a trait of rationality, and that some people have more of it than others. Doesn't that undercut the contrast between acting from a trait and acting rationally? The answer is no: the distinction I had in mind is one between choosing how to act when only one course of action is rationally permissible (hence, we neither need nor can appeal to a trait in order to explain why a person does what she does—she does what is simply rational) and choosing one of several rationally permissible alternatives. But there is a difference between taking the only rationally permissible course of action available and taking the most rational among several possible alternatives, and an appeal to a trait of rationality will be appropriate in the latter but not in the former case. As with other traits, we ascribe the trait of rationality to people who choose one of several rationally permissible courses of action; only this time, what makes that course distinctive is not that it is "loyal" or "honest," etc., but that it is the most rational. I thank Michael Huemer for pressing me to explain this point.

29. It is, of course, possible that there is some highly improbable but still rational explanation of Abe's action. For instance, imagine the ID in the wallet is a business company ID card, and that Abe has just been fired from the company in question. In addition, Abe has recently had a bad car accident, and the prospect of having to look at the driver's license in the wallet occasions grim thoughts in him. If that is the case, it may be rational for Abe to throw his own wallet in the trash (indeed, it may be that the reason he has lost the wallet in the first place can be traced to an unconscious desire to get rid of it). I am assuming here that such an improbable but rational explanation is unavailable, and that the only rational thing for Abe to do is to take back the wallet.

30. This explanation may, at a deeper level, be underwritten by a traits explanation; for instance, Abe may be a womanizer who would never miss the chance to meet a beautiful young woman. But it may also be that Abe is not a womanizer; he simply wants to return

the wallet to this particular woman. Similarly, the good Prince Myshkin I mentioned in the beginning is described by Dostoyevsky as falling in love with and making a decision to find a woman—Anastasia Filipovna—upon seeing her photo, though Myshkin is, by no means, a womanizer.

31. Of course, one could say that being fickle is a character trait, and in some sense, that is true, but unlike other traits, it has little predictive power. The only thing it can tell us is that the agent can be expected to behave unpredictably.

32. I mention in passing that traits may lack a behavioral aspect for reasons other than a deliberate attempt at masking a trait on the part of the agent possessing it. For instance, it could be that a person is disposed to be kind but, having chosen the life of a monk in seclusion, never actually does anything kind.

33. Not all values underwrite character traits. Nonmoral values, such as valuing antique objects may not, but the motivational aspect of character traits has to do with values.

34. Note that a person may acquire the false belief that she values something when her desire is really pathological, for instance, a person with OCD may falsely believe she is just scrupulous. See Walter Sinnott-Armstrong and Jesse Summer's chapter in this volume. There should be no particular puzzle about this: values cannot be equated with what a person believes she values, since we can be mistaken about what we value, as when the powermonger wrongly believes she is motivated by a sense of duty.

35. He resorts to a kind of soft determinism.

36. The following issue needs separate discussion, but I wish to note that psychopaths, on my account, are not morally responsible for their characters if they are not capable of the sort of rational reflection that could lead a person to change. And evidence suggests that they are not capable: to my knowledge, there are no documented cases in which a clinical psychopath has had a moral conversion, though there are many such cases involving non-psychopathic people, including people who have a history of violent behavior, who have experienced a moral conversion.

37. I and my co-author Jon Tresan argue at length against determinism in "Will Retributivism Die and Will Neuroscience Kill It?" (2015).

38. Of course, the propensity to reflect may itself constitute one particular character trait: being reflective. But first, this is just one possible character trait among many. Second, and more important, as with rationality, which is both a universal trait and may be a character trait as well, the ability to reflect, while it may be a trait, is also an ability universal to rational human beings, and it provides a basis for our responsibility for our characters.

39. This possibility points to a limitation of the virtue theory idea that good actions result from a virtuous nature. It is, of course, possible to argue that if a person must overcome a morally bad emotion in order to perform a good action, she is less praiseworthy. Rosalind Hursthouse (1999) defends this view. Nomy Arpaly and Timothy Schroeder (2014, 190) oppose it.

40. Thus, Ecclesiastes 1:18 reads, "With much wisdom comes much sorrow."

41. For instance, a person who has previously been told that his jokes are inappropriate and who, through no particular fault of his own, fails to be humorous, can simply abstain from making jokes.

42. On the role of historical circumstances on character, see Daniel Little's "Character and History," this volume.

43. As mentioned in the beginning, however, we often treat people's character traits as though they are deterministic mechanisms. The reason for this is that, for practical purposes, we are primarily concerned with what people are likely to do, not what they are capable of doing upon reflection.

44. Regarding the question of how such actions could still count as reason-based, see Fileva 2013.

References

Arpaly, Nomy, and Timothy Schroeder. 2014. *In Praise of Desire*. Oxford: Oxford University Press.

Brandt, Richard. 1970. "Traits of Character: A Conceptual Analysis." *American Philosophical Quarterly* 7: 23–37.

Brandt, Richard. 1992. *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.

Brody, Nathan. 1988. *Personality: In Search of Individuality*. New York: Academic.

Buss, Sarah. 2012. "Autonomous Action: Self-Determination in the Passive Mode." *Ethics* 122: 647–691.

Creemers, Daan, R. H. Scholte, R. C. Engels, M. J. Prinstein, and R. W. Wiers. 2012. "Implicit and Explicit Self-Esteem as Concurrent Predictors of Suicidal Ideation, Depressive Symptoms, and Loneliness." *Journal of Behavior Therapy and Experimental Psychiatry* 43: 638–646.

Doris, John. 2002. *Lack of Character*. Cambridge: Cambridge University Press.

Fileva, Iskra. 2008. "The Neutrality of Rightness and the Indexicality of Goodness." *Ratio* 21: 273–285.

Fileva, Iskra. 2013. "A Puzzle about Knowledge in Action." *Logique et Analyse* 56: 287–301.

Fileva, Iskra, and Jon Tresan. 2015. "Will Retributivism Die and Will Neuroscience Kill It?" In "Philosophical Approaches to Social Neuroscience," special issue, *Cognitive Systems Research* 34–35: 54–70.

Fleeson, William. 2004. "Moving Personality beyond the Person Situation Debate." *Current Directions in Psychological Science* 13: 83–87.

Funder, David. 1994. "Explaining Traits." *Psychological Inquiry* 5: 125–127.

Gilbert, Margaret. 2006. "Character, Essence, Action: Considerations on Character Traits after Sartre." *The Pluralist* 1: 40–52.

Hill, Amelia. 2009. "Healthy Food Obsession Sparks Rise in New Eating Disorder." *The Guardian*, http://www.theguardian.com/society/2009/aug/16/orthorexia-mental-health-eating-disorder.

Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

Korsgaard, Christine. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Lucas, Richard, and Brent Donnellan. 2009. "If the Person-Situation Debate Is Really Over, Why Does It Still Generate So Much Negative Affect?" *Journal of Research in Personality* 43: 146–149.

Mele, Alfred. 2004. "Volitional Disorder and Addiction." In *The Philosophy of Psychiatry*, ed. Jennifer Radden, 78–88. Oxford: Oxford University Press.

Mischel, Walter. 2009. "From *Personality and Assessment* (1968) to Personality Science, 2009." *Journal of Research in Personality* 43: 282–290.

Mischel, Walter, Yuichi Shoda, and Rodol Denton. 2002. "Situation-Behavior Profiles as a Locus of Consistency in Personality." *Current Directions in Psychological Science* 11: 50–54.

Moore, Adrian. 2003. *Noble in Reason, Infinite in Faculty: Themes and Variations in Kant's Moral and Religious Philosophy*. London: Routledge.

Moreau, Sophia. 2005. "Reasons and Character." *Ethics* 115: 272–305.

Ryle, Gilbert. 1949. *The Concept of Mind*. Chicago: University of Chicago Press.

Taylor, Craig. 2001. "Moral Incapacity and Huckleberry Finn." *Ratio* 14: 56–67.

Williams, Bernard. 1982. "Persons, Character and Morality." In *Moral Luck*, 1–19. Cambridge: Cambridge University Press.

Williams, Bernard. 1993. "Moral Incapacity." *Proceedings of the Aristotelian Society* 93: 59–70.

Wilson, Timothy. 2004. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Belknap Press.