Social Psychology, Moral Character, and Moral Fallibility*

LORRAINE BESSER-JONES
University of Waterloo

In recent years, there has been considerable debate in the literature concerning the existence of moral character. One lesson we should take away from these debates is that the concept of character, and the role it plays in guiding our actions, is far more complex than most of us initially took it to be. Just as Gilbert Harman, for example, makes a serious mistake in insisting, plain and simply, that there is *no* such thing as character, defenders of character also make a mistake to the extent that they imply there is *no* problem raised by the psychological literature for either the concept of character or the nature of character-based ethics. My hope for this paper is to avoid both of these mistakes by first, exploring exactly what is the concept of character that is so firmly rooted in our philosophical and everyday thinking; and second, exploring the implications of the psychological literature for this appropriately understood concept of character. In so doing, I will come to a resolution that vindicates the existence of character, while at the same time calls attention to the real and serious problem suggested by the psychological evidence. This, we will see, is a problem of moral motivation.

"*A practically relevant character ethics should have something to say about securing ethically desirable behavior*"[1]

The concept of a person's character is one engrained in our culture that plays an important role in some of our most fundamental deci-

---

[1] Doris 2002, p. 110.

sions: we cast votes for politicians based largely on our beliefs about their characters, we hire and fire people on the basis of their characters, we choose our life partners with an eye to their characters, and we raise our children with the goal of instilling in them good characters. All of this seems like common sense: a person's character is what makes her who she is and, many think, is what lies at the core of her very self.

Despite the fundamental role the concept of character plays in our everyday lives, the concept itself remains something rather elusive. Amongst philosophers, the most common way of understanding "character" is, following Aristotle, in terms of a settled disposition to act in certain ways. Thus character traits have been taken essentially to be dispositional traits: the honest person possesses the disposition to tell the truth; the just person possesses the disposition to act justly, and so on.

In recent years, however, this understanding of character has come under fire, most notably through the works of John Doris and Gilbert Harman.[2] Doris' book, *Lack of Character*, lies at the forefront of this attack, giving a detailed and sustained argument that challenges this standard view of character and, in particular, the role most take it to play in motivating people to act. Drawing on empirical psychology and specifically the doctrine of situationism, Doris argues that people lack character traits, where "character" is understood to consist of the dispositional traits that lead people to act in distinctive and predictable ways. This conclusion—if sustainable—has dramatic implications for moral theories, many of which, particularly Aristotelian virtue theories, depend to some extent on the idea that people possess characters that influence their behavior in meaningful ways. Thus, Doris argues, for example, that since social psychology shows the Aristotelian psychology to be "descriptively inadequate," this limits its ability to have "something to say about practical questions"—an ability that Doris rightly thinks to be critical to any normative theory.[3]

Doris' and Harman's call for such radical changes has met with spirited resistance: defenders of character have argued that Doris' and Harman's interpretation of the empirical evidence is misguided, that their interpretation of Aristotle is unwarranted, that their conclusions are too extreme.[4] Most of these objections have substantial merit: some of the experiments do seem to allow for alternate explanations, and there do seem to be various ways for the Aristotelian to escape Doris' charge of descriptive inadequacy. The message we get from Doris'

---

[2]   See Doris 1998, 2002 and Harman 1999, 2000, 2003.

[3]   Doris 1998, p. 505 and 513, respectively.

[4]   See, for example: Athanassoulis 1999, Kamtekar 2004, Kupperman 2001, Merritt 2000, Miller 2003, Solomon 2003, Sreenivasan 2002, Swanton 2003, Vranas 2005.

and Harman's opposition is one of skepticism and hope: since the psychological evidence does not provide knock-down evidence against our attributions of character, we have reason to be skeptical of implications drawn from it, and, given the fundamental role that we make of character attributions, hopeful that there is something meaningful that underlies these attributions.

This position of skepticism and hope does not, however, present a satisfactory resolution to the debate over character. It is fair to say that the one lesson we can, and should, take away from this debate is that the concept of character, and the role it plays in guiding our actions, is far more complex than most of us initially took it to be. Just as Harman, for example, makes a serious mistake in insisting, plain and simply, that there is *no* such thing as character, defenders of character also make a mistake to the extent that they imply there is *no* problem raised by the psychological literature for either the concept of character or the nature of character-based ethics.[5] My hope for this paper is to avoid both of these mistakes by first, exploring exactly what is the concept of character that is so firmly rooted in our philosophical and everyday thinking; and second, exploring the implications of the psychological literature for this appropriately understood concept of character. I argue that character consists not solely in behavioral dispositions, but also in one's moral commitments, and the ways in which these commitments interact with one's dispositions. I go on to argue that the real and serious problem suggested by the psychological evidence is a problem of moral motivation. I conclude by offering some suggestions, drawn from empirical research on self-regulation, about how moral theorists might works toward solving this problem.

§1. Moral Character

Essential to the success of Doris' and Harman's argument that people "lack character" is that they are working with a correct account of what, in fact, character is. If they are not, then their arguments miss their target.

According to both Doris and Harman, a character trait is essentially a dispositional trait that leads an agent to act in trait relevant fashions across a wide range of circumstances. These sorts of traits are the "robust" character traits that the psychological experiments call into question: the experiments show that how a person tends to act depends more on situational factors than on anything internal to the

---

5   Doris, to his credit, recognizes part of the complexity involved in the very concept of character, as seen through his attempt to give content to the idea of "local" character traits.

person. Since people fail to exhibit robust dispositions, if character traits are essentially these robust dispositional traits, then people must lack character traits. As some philosophers are beginning to note, however, this conception of character (consisting in the possession of robust character traits) departs from both philosophical and ordinary usage of the concept. Christine Swanton and Rachana Kamtekar, for example, argue that the account of character presupposed by Doris and Harman is insufficient, and in particular, does not conform to a traditional, Aristotelian account of character. Character traits are not simply dispositional traits proved to exist or not through people's behavior; rather there is something more to character traits and to people's characters. The difficulty is explaining exactly what this "something more" is.

Swanton describes the "something more" aspect of character in terms of "fine inner states." She argues, "for a virtue ethicist, a virtue is not simply a disposition to perform acts of a certain type (e.g. beneficent acts). The possession of virtue requires also the possession of fine inner states."[6] Unfortunately, Swanton does not go on to detail explicitly exactly what these "fine inner states" are, but presumably they are wrapped up in the traditional Aristotelian conception of a virtue, which, in addition to the disposition to act, involves also dispositions to think, feel, and judge in certain ways.

Kamtekar provides a fuller account of the "something more" aspect of character. She criticizes Doris and Harman for "isolating [character] from how people reason," and argues that one's character is inherently connected not just to dispositions to act, but also to the ways in which people deliberate about how to act.[7] She writes: "the conception of character in virtue ethics is holistic and inclusive of how we reason: it is a person's character as a whole (rather than isolated character traits), that explains her actions, and this character is a more-or-less consistent, more-or-less integrated, set of motivations, including the person's desires, beliefs about the world, and ultimate goals and values."[8]

While Kamtekar later goes on to imply that character itself does not *include* the desires, beliefs, goals, and so on, except insofar as these factor into our "dispositions to act and feel in particular ways in response to rational considerations", in highlighting the integral role character plays in reasoning as well as acting, Kamtekar makes significant progress towards identifying what Doris and Harman overlook in their

---

[6]   Swanton 2003, p. 30.

[7]   Kamtekar 2004, p. 460.

[8]   Kamtekar 2004, p. 460.

critique of character.[9] Doris' and Harman's version of character simply places too much importance on dispositions to act.

While in Doris' official statement of what character traits are, he is careful to include "other things", writing, "as I understand it, to attribute a character or personality trait is to say, *among other things*, that someone is disposed to behave a certain way in certain eliciting conditions,"[10] after making this qualification, he proceeds in his discussion to ignore the other facets of character, and focuses primarily on behavioral dispositions. This emphasis can be seen through his understanding of the role of character attributions, which he takes to have an essentially predictive purpose: we attribute them on the basis of "behavioral consistency" and we invoke them as means to predict and explain people's behavior.[11]

Certainly what Doris says here is partly true, but it does not, I think, reflect the whole nature and role of the concept of character. Most significantly, it under-appreciates the evaluative component of character attributions.[12] After all, one important role the concept of character plays is to give us insight into the person who possesses it: not just about what she will do under certain circumstances, but about what she believes and holds to be important. These are aspects of character that affect the evaluative aspect of character attributions. Talk of character—especially moral character—involves an irreducible evaluative component: it involves categorization of the person herself, who we see as represented through her character.

Interestingly, Doris both recognizes this evaluative aspect of character traits, and thinks it is what differentiates character traits from personality traits such as shyness. He writes that "character" usually involves an evaluative component that "personality" lacks, citing as an example the trait of honesty. "The honest person," Doris argues, "presumably behaves as she does because she values forthrightness, while the introvert may not value, and may in fact disvalue, retiring behavior in social situations."[13] Despite these evaluative elements that are attached to "character" traits, and not typically "personality" traits, Doris insists that both share common ground on a more fundamental level, having in common "behavioral consistency as a primary criterion

---

[9]   Kamtekar 2004, p. 479.

[10]   Doris 2002, p. 15, my emphasis.

[11]   Doris 1998, p. 18 and 15.

[12]   See Kamtekar 2004, p. 478-479 for a helpful discussion of this evaluative aspect.

[13]   Doris 2002, p. 19.

of attribution'' and thus, in his critique of character, treats character along the same lines as personality.[14]

It is at this point that I think Doris makes a crucial mistake. The evaluative component Doris addresses is, I shall argue, a distinctive feature of *moral character*; understanding this role not only allows us to see that character involves more than dispositions, but also helps us to understand why we often take dispositions to be a central (albeit not comprehensive) aspect of character.

My dog licks my face every morning when the alarm goes off and barks every time the doorbell rings. I know these things about her dispositions, but most of us would agree that this knowledge of my dog's dispositions is very different than the knowledge we have about people's characters, and the use we make of our knowledge of people's dispositions.

What use do we make of our knowledge of people's dispositions, and why is this knowledge important? We use our knowledge of people's dispositions to tell us something *further* about the person (i.e., something beyond "they are disposed to X given circumstances A, B, C"), and this is a significant reason why we take such knowledge to be important. We find explication of this sort of view in Hume, who argues that acts are important only insofar as they give us knowledge of the person's "principles" of mind and temper:

> 'Tis evident that when we praise any actions, we regard only the motives that produc'd them, and consider the actions as signs or indications of certain principles in the mind and tempter. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still consider'd as signs; and the ultimate object of our praise and approbation is the motive, that produc'd them.[15]

Hume's claim helps to explain the difference between our reactions to my dog's dispositions and our reactions to a person's dispositions: in the latter case, we think the dispositions are important partially insofar as they tell us something about the beliefs, motives, and values of the person, whereas in the former case, the most we can learn from the dispositions is the extent to which my dog has learned to react to certain stimulus. We do not presuppose, except in moments of fancy, that my dog's disposition to lick my face every morning results from her beliefs about how best to express her love for me. But, when a

---

[14]   Doris 2002, p. 20.

[15]   Hume, T 3.2.1.2 (p. 307).

person repeatedly displays a particular disposition, we presuppose that her disposition results from certain beliefs she has about how to act.[16] This is likely a significant reason why we tend to take dispositions as *the* central feature of character: because, as Hume says, we consider them to be "external signs" of internal states that are hard to identify, yet truly lie at the core of one's character.

Thus, while Doris and Harman are right to take an agent's dispositions to be reflective of her character, they fail to acknowledge *why* we do so, and this leaves them with a very watered-down account of character that equates the nature of a person's dispositions with that of an animal's dispositions. And, more importantly, this leaves them with an account of character that is very different than the one employed in everyday discussions of character.

This very same reasoning applies not only to the behavioral dispositions Doris and Harman take to be distinctive of character, but also to the "dispositions to respond" that neo-Aristotelians, such as Kamtekar, take to be definitive of character. Kamtekar, recall, criticizes Doris and Harman for focusing too narrowly on behavioral dispositions and so for failing to recognize the central role character plays in an agent's reasoning process. She argues that central to the virtue ethical concept of character found in Aristotle, among others, are dispositions to respond in certain sorts of ways: "a virtuous disposition is a disposition to act and feel in particular ways in response to rational considerations; it is expressed in our decisions, which are determined through rational deliberation."[17] I grant that these dispositions are important facets of character; but again, let us not neglect *why* we take them to be important facets of character: because we assume that they reflect corresponding beliefs of the agent. We assume that when an agent judges she should X, and feels guilty when she fails to X, that this is because she *believes* that X is the appropriate action in these circumstances. While often we may be wrong in making these assumptions, it is nonetheless the case that we take such dispositions to be important largely because we take them to reflect something about the person herself, and not just about the extent to which she has developed certain capacities to respond to particular stimuli. The beliefs that give rise to dispositions to feel and judge—as well as to act—are essential to evaluations of character,

---

[16]   Now, as Doris and Harman have shown us, very often this presupposition is mistaken. I will address the implications of this finding later in this paper; my point here is only that we think dispositions are important largely insofar as we take them to be reflective of something more about the person. Thus any account of character should not focus solely on dispositions.

[17]   Kamtekar 2004, p. 479.

and should be recognized as themselves being part of a person's character.

After all, most people take it as a given that knowledge of dispositions alone cannot give us a full picture of a person's character. Two people can share the same dispositions, but nonetheless differ in their characters. Take the case of Jane and Mark. Jane believes that everyone deserves equal treatment, and condemns racism. Yet, whenever she sees a black man walking down the street, she averts her eyes and, if possible, crosses the street. She feels guilty when she does so, but nonetheless cannot help herself from acting in these ways. Mark, on the other hand, holds racist beliefs about the inferiority of African Americans. He, too, averts his eyes and crosses the street when encountering a black man. He, too, feels guilty when he does so, after all, it is not as if he holds anything "personal" against the man, but is simply acting on the basis of what he thinks is right.

Jane and Mark share very similar dispositions—both behavioral dispositions, and dispositions to feel certain ways in response to their behavioral dispositions—nonetheless, all would agree that they have different moral characters: Jane is what psychologists label an "aversive racist"—one who is "consciously non-prejudiced yet unconsciously prejudiced";[18] her character is significantly different than Mark's, whose character is prejudiced on all levels. To claim that they have the same character in virtue of having the same dispositions would be a serious mistake for it cuts us off from a wealth of explanatory tools, such as making sense of the distinction between aversive racists and individuals who are both implicitly and explicitly racists. This suggests that more than dispositions constitute people's characters, be they behavioral dispositions or dispositions to feel and judge.

§2. A Revised, Complex Account of Character

From this discussion, we can piece together an initial account of moral character relatively easily. I propose that moral character consists of the following components:

(a) The moral beliefs to which one is evaluatively committed, particularly those relating to an agent's interactions with others;

---

[18]   Son Hing, L.S, et al 2005, p. 275.

(b)   An agent's dispositions to act; and

(c)   The nature and degree to which one's moral commitments influence one's behavioral dispositions. This includes beliefs one has about how one should act, one's dispositions to feel and judge in certain ways, as well as one's desires and/or intentions to act in certain ways.

Each of these parts makes up the entire package of one's character, which is robust and complicated, and not represented by any single component. Just as, I have argued, it is a mistake to identify "character" with dispositions, it is also a mistake to identify "character" with beliefs; rather, "character" is both of these things, and the ways in which they interact. Let me now explain each of the components of this account of character in detail.

§2.1 Part (a): Moral Commitments

Part (a) consists of those evaluative commitments that are central to moral character; in short, those that comprise one's beliefs about morality. Some people are committed to the principles that lying is wrong, and that it is right to help others when you can. Others commit themselves to the Golden Rule. Some people have more localized moral commitments: lying is wrong, except when it protects your family; it is important to give blood; victims of natural disasters deserve charity; and so on.

To be morally committed to a particular belief is to be committed to believing that it is true. The sort of moral commitment requisite here is characterized by a resistance to revision of these beliefs in the face of evidence for competing beliefs. While we are not inflexibly committed to them, we nonetheless expect and believe that our moral commitments will hold up under normative cross-examination. This component of character, it is worth emphasizing, has an epistemic status, although the beliefs that comprise it have a practical subject matter. Moral commitments contain beliefs about the rightness and wrongness of certain behaviors; beliefs that are quite possibly independent of an agent's understanding of what *she* should do, and beliefs which may or may not give rise to dispositions to act in accordance with those beliefs.

Although it is tempting to think that agents who affirm a moral belief will consistently act in accordance with said belief, let us not forget that human nature is flawed: human beings are, quite simply, morally fallible. This is something most of us are quite willing to accept as a fact

about ourselves.[19] For some reason, however, we are less apt to recognize this very same moral fallibility in others; hence, when someone professes to believe it is wrong to lie, and proceeds to lie, our first instinct is to question the person's sincerity. We find it hard to believe that a person can *really* believe something is wrong to do, and not act on that belief. Thus, when evaluating this situation in other people, we tend to think to think that, in cases where beliefs and actions come apart, these are cases of insincerity and hypocrisy, as opposed to moral fallibility.

Why the disparity between our self- and other-evaluations? The answer is clear: while we can be fairly confident in our knowledge of our own beliefs, we have limited access to the beliefs of others. Very often, the only way in which we can have access to another's beliefs is through observation of her actions; thus, when a person claims to believe something and acts in a way that conflicts with this belief, we take the *action* to be stronger evidence that the person's testament of her beliefs. But why should we take actions to consist in the stronger evidence? Why are deeds taken to be stronger than words? The mere fact that we have limited epistemic access to a person's beliefs should not entail that we should place all of our faith in a person's actions.[20]

---

[19]   Tom Regan, for example, describes what I take to be a common phenomenon, writing:

> We all sometimes fail to live up to our professed standards. For example, although I value truth, I have lied on occasion. I have sometimes said that I had done something when I knew full well that I had not, and I have sometimes said that I had not done something when I knew full well that I had. I doubt that I am exceptional . . . Does this mean that I am being hypocritical when I say that I value truth? I do not think so. What the conflict between my values and my moral behavior illustrates is my moral fallibility. For a variety of reasons (selfishness is one, and weakness of will is another), I sometimes fail to honor my values in my life. Sometimes my actions contradict values in which I really believe'' (Regan 2001, p. 168).

Regan's words ring true for most of us: through our own experiences, we know *believing* something is valuable and right-to-do is often not enough to successfully motivate us to act on said belief.

[20]   At this point, a standard move is to look beyond behavioral dispositions and to a person's dispositions to feel and judge in certain ways, particular in response to her actions. Many people feel more comfortable attributing sincerity to a person who feels remorse when she fails to act on her moral commitments. I'm not sure what real work this does, for, at the end of the day, (1) just as we have to take a person's word for what she believes, we have to take a person's word for what she claims to feel; and (2) just as there are factors that prevent people from exhibiting behavioral dispositions, there are factors that can prevent a person from exhibiting dispositions to feel and judge. Thus it is not going to be sufficient to claim that the move to these other sorts of dispositions can give us *true* access to a person's beliefs, although I agree that, other things equal, these dispositions do help us to give insight into a person's beliefs.

Given that we have all experienced cases of moral fallibility, we ought not to be so quick to make the charge of hypocrisy or insincerity in the case of others. While under some circumstances this charge may be warranted (for example, if a person professed to value honesty but *always* lied), before making this charge we need to rule out the factors—such as *akrasia*—that can make us liable to moral fallibility, and that can create a wide gap between evaluation and motivation.[21]

Michael Smith rightly takes the existence of this gap to be "commonplace, a fact of ordinary moral experience."[22] Empirical research supports Smith's claim that incidences of *akrasia*, self-deception, *accidie*, and hypocrisy, are common and not just reserved for a special group of morally- or psychologically-defective people. To illustrate this point, consider the significant amount of empirical research testing the extent to which health beliefs such as "smoking is bad" affect health behavior.[23] As Abraham et al report, numerous studies show that the "perceived severity of health risks (e.g., 'It would be very bad to contract skin cancer.') . . . [is] a weak correlate of health behaviour."[24] One such study explores the connection between women's beliefs about abortion and actual use of contraceptive devices, and found that there was very little, if any, connection: the study found no difference in normative beliefs about abortion between those women who consistently use contraception and those who use contraception only inconsistently.[25]

My guess is that most readers will not find this study, or others like it, very surprising. Whether in the case of health beliefs or moral beliefs, it is very common for beliefs and actions to become disparate. As a result, we cannot draw justified conclusions about the content of an agent's moral (or health-related) commitments on the basis of her behavioral dispositions. Nonetheless, while it is difficult to ascertain what people's beliefs are, this problem of knowledge should not

---

[21] Other factors include self-deception, which prevents the agent from recognizing that, even though it is all-things-considered right to tell the truth in circumstances X, Y, Z, she herself is in these circumstances, so should tell the truth, as well as *accidie*, cases in which a person generally lacks the desire to act on what she values.

[22] Smith 1994, p. 120-121.

[23] It is important to distinguish here between beliefs reporting empirical or theoretical claims, such as "smoking is bad for me" and "it is right to help people" from beliefs concerning behavioral factors such as "I can do this." Self-efficacy beliefs, of the latter sort, have been shown to play a significant role in motivation, whereas the former sort of beliefs has been shown to have little role in motivation (Abraham et al., 1998).

[24] Abraham et al., 1998, p. 575.

[25] Muchow 1987, p. 64.

preclude us from believing that people's beliefs do constitute a substantial part of their character, be it a sometimes opaque part.

I have argued that an agent's moral commitments are an important part of her character, and that, given moral fallibility, such moral commitments may not be exhibited by her behavioral dispositions; let us now turn to a brief discussion of the nature of these dispositions.[26]

§2.2 Part (b): Dispositions

Dispositions are what most philosophers tend to think of when they think of "character" and so I will not say much about them here, except to emphasize their practical nature.

While people have dispositions to do a great many things, such as to feel certain ways, or to form certain sorts of judgments, the dispositions I want to focus on here are dispositions to behave in certain sorts of ways.[27] Attributions of behavioral dispositions are solely practical and describe the ways in which people tend to act without making explicit reference to their moral commitments.[28] People have dispositions to tell the truth, to keep promises, to lie, or to break promises, in the relevant situations. These dispositions can be understood and described without any reference to one's moral commitments regarding the rightness or wrongness of telling the truth, or of keeping promises.

In some cases, one's behavioral dispositions reflect one's moral commitments, and we tend to assume that they do. When an agent exhibits a truth-telling disposition, we assume her disposition reflects a moral commitment such as "It is wrong to lie to people" or "lying treats people wrongly." In fact, we would be surprised to learn that the agent who always tells the truth, and so exhibits a truth-telling disposition, held no corresponding moral commitments about the wrongness of lying. While such a situation may be surprising, let us not question its possibility: as the preceding section has shown, very often there are factors that make it the case that one's behavioral dispositions and corresponding actions are not necessarily reflective of one's moral commitments.

---

[26]   My discussion in this section clearly commits me to a denial of some versions of motivational internalism, and in particular, any version of internalism that denies the possibility of incidences of moral fallibility such as akrasia.

[27]   I do think that other dispositions, namely dispositions to feel and judge, are important facets of character, but think that these arise primarily on reflection of the interaction between moral beliefs and behavioral dispositions, so have grouped them in the third component of character.

[28]   Although, as I have argued, we tend to take people's behavioral dispositions to be reflective of their beliefs.

§2.3 Part (c): Interactions

While so far we have focused on the fact that behavioral dispositions can conflict with moral commitments, there are obviously a number of other ways in which they can and do interact, and the extent and nature of these interactions make up the third component of my proposed account of character. While here is not the place to identify all the possible ways moral commitments can interact with dispositions, let me identify two of the most salient features we should look to for evidence of the nature and degree of their interaction.

The first is obvious: what is the correlation between the two? Do the agent's behavioral dispositions reflect her moral commitments always, sometimes, most of the time, or never at all?

The second looks beyond the mere correlation, and explores the agent's practical attitudes. How does the agent believe she should act? What sort of person does she want to be? Does she think she should act on her moral commitments? Does she feel guilty when she does not? Answers to these sorts of questions give us important information as to the nature of her character. This second feature clearly informs the latter. The content of one's practical attitudes has a clear connection to the nature of the correlation between the agent's moral commitments and dispositions.

Notice that it is answers to these questions that will give us perhaps the most substantive insight into a person's character: they will tell us whether she has integrity, whether she is weak-willed, whether she suffers from alienation, and so on. These are important features of characters, as well of our evaluations of characters, and are features that can only be learned through examination of the ways in which our moral commitments affect and engage with our behavioral dispositions.

The concept of character, thus, is much more complex than those accounts of character that interpret it on purely dispositional grounds recognize it to be. We cannot evaluate a person's character on the basis of dispositions alone. To do so is to cut ourselves off from a wealth of knowledge about the *person* herself. And, ultimately, when it comes to character appraisal and attribution, what we are really trying to do is not simply to predict how people are going to act or respond to certain situations, but is to gain insight into the person. We can only gain this insight if we look beyond the dispositions a person exhibits and into the beliefs and practical attitudes she holds.

§2.4: Why this understanding of character

I hope to have already provided an intuitive account of why this understanding of character, which I will refer to as the "complex account of

character'', is preferable to one that focuses primarily on dispositions. Let me now make the case more formally by exploring the ways in which this account of character underlies common usage of the concept. First, however, a caveat: in suggesting that "character" comprises the three parts detailed above, I do not claim that this is an exhaustive list of the components of character. There very well may be further nuances to our understanding of character not addressed here; nonetheless, I think the three parts I have identified capture the primary and most central components of character.

While when we speak of "character *traits*" we tend to think of behavioral dispositions, talk of character is not limited to talk of character traits. There is a distinctive set of evaluative labels that invoke the concept of character understood not as a mass of specific dispositions but rather as a complex whole, involving the components I have outlined.

Consider first labels of "strong" characters and "weak" characters, as well as more fine-grained classifications such as "strong but bad" characters and "weak but good" characters. Hitler, we might say, has a "strong but bad" character, while the akratic is often characterized as having a "weak but good" character. The strength components clearly refer to the extent to which an agent is able to act on her moral beliefs, while the pure evaluations of "good" and "bad" refer to the nature of those moral beliefs. Hitler has a set of "bad" moral beliefs, yet has forged a strong (consistent, reliable) connection between those beliefs and his dispositions, while the akratic (at least as is typically discussed in the philosophical literature) likely has a set of "good" moral beliefs, yet has been unable to form a strong connection between her beliefs and her dispositions.[29]

Consider further the concept of "integrity", a label that most certainly refers to a person's character: to claim that a person has or lacks integrity is to make a claim about the very nature and structure of her character. Integrity concerns the nature of one's beliefs (do they point to a singular moral purpose?), the relationship between those beliefs (are they consistent?), as well as the relationship between one's beliefs and desires (do we desire to have these beliefs, and vice versa?). While, on most accounts, integrity does not function as a typical character "trait" such as honesty, it nonetheless is a mode of character evaluation, which again suggests that in evaluating characters as wholes, we look at much more than dispositions, and take beliefs to be a central component of character.

---

[29]   I am using "good" and "bad" here only to appeal to regular usage of these phrases, without making any commitments as to how we should go about evaluating characters.

Finally, consider the popular evaluative phrase "acting out of character". When invoked, this phrase typically refers to incidences in which someone displays dispositions that conflict with his or her moral commitments: the person who cheats on her taxes but otherwise embraces a strong sense of civic commitment acts "out of character". While I suppose there is a sense in which an account of character that takes character to consist primarily in dispositions can make sense of this phrase, by understanding "out of character" to refer simply to incidences in which agents exhibit dispositions that conflict with their usual dispositions, the complex account of character offered here provides a reading of the phrase that gives us significantly more insight into the situation.

§3. Revisiting the Empirical Evidence

Once we recognize that one's character consists in more than behavioral dispositions, we see that the only reason why Doris and Harman reach the conclusion that people lack character is because they are working with the wrong account of character. Clearly, once we recognize that character extends beyond one's robust dispositional traits, we see that we simply cannot make claims about the very *existence* of character on the basis of observations of a person's behavior. But what can we learn about a person's character on the basis of such observations?

If people do lack robust dispositional traits (again, these are traits that lead them to act in a similar fashion across a wide range of trait-relevant circumstances), then this is certainly going to impact our understandings of people's characters. Thus, we need to examine the extent to which the evidence proves what Doris and Harman take it to prove, with respect to robust dispositional traits.

Many people have argued that the empirical evidence simply does not support a denial of the existence of robust dispositional traits. Nafsika Athanassoulis, for example, argues that there are alternative interpretations of the experiments taken to be central by Harman, in particular. In response to the Milgram experiments, where subjects were instructed to administer shocks to people who provided incorrect answers to a series of questions, she argues that the only conclusion to be drawn from the fact that almost all of the subjects administered shocks past the point of severe danger is the hypothesis that the administrators were seeking to test, and designed the experiment to test: that "most people would follow orders", a conclusion very different from "the conclusion that people possess no character traits."[30]

---

[30]   Athanassoulis 2000 p. 216.

Gopal Sreenivasan takes a similar approach to challenging Doris' and Harman's interpretation of the empirical data. Like Athanassoulis, he questions the legitimacy of drawing conclusions from experiments designed to test alternative hypotheses, but he also worries about the methods employed by those specifically testing situational influences on character. The experiments favored by situationists are ones that rely on objective determination of trait-existence, rather than subjective reporting of trait-existence. Situationists avoid subjective reports on the suspicion that people make the fundamental attribution error to themselves as well as to others; however, as Sreenivasan notes, when we examine the situation from an objective standpoint, we ignore the "subject's own construal of the situation."[31] A subject may be doing what *she* thinks is the compassionate act, but we, as objective observers, might determine otherwise. A second problem Sreenivasan raises concerns how we determine what sorts of behavior are truly relevant to the trait in question: is the best way to test for compassion putting a coughing man on the side of the sidewalk and seeing what students who are hurrying to an appointment will do?[32] Sreenivasan thinks the answer is no, and he is probably right.

We do have reason to be suspicious of Doris' and Harman's blanket claim that people lack robust dispositional traits, but this does not mean that we have nothing to learn from their arguments and the experiments they drawn on. One thing we can be certain of is that the large majority of people fail to do what is right in the experiments in question. Doris' book contains numerous experiments in which large groups of people let someone suffer, make someone suffer, fail to get help, take what is not theirs, and so on. Even if these were isolated acts, they still point to a big problem that warrants serious attention.

There are two possible explanations for this behavior: The first is that people have the ''wrong'' sort of moral beliefs: for example, they think it is all-things-considered right to follow authority figures, even when these authority figures order them to harm others. The second is that people have the "right" sort of moral beliefs, yet—for whatever reason—are not acting on them. Moral theorists have a tendency to think that the first sort of cases are their primary concern: that is what is most important is figuring out and teaching others the "right" moral beliefs. The assumption seems to be that, once people have the right moral beliefs, they will then (a) know exactly what to do in any

---

[31]   Sreenivasan 2002, p. 47.

[32]   This example refers to the "Good Samaritan" experiment conducted at the Princeton Theological Seminary. For discussion of this, see Darley and Batson 1973, as well as Doris' discussion in Doris 2002, p. 33-34, and Harman 1999, p. 323-324.

situation; and (b) actually do the right thing. It is this assumption that I think the empirical evidence shows to be false.

While the empirical evidence most often does not tell us anything about the content of a subject's moral beliefs, we can presume that most people believe certain things about their conduct. For example, it is a safe assumption that many, if not most, of the subjects in the Milgram experiments *believed* that it is wrong to inflict pain on others. While we can argue about the exact percentages of people who held such beliefs, it is safe to say that a large percentage of them held basic moral beliefs of this sort, yet did not act on them.

Putting this problem into the terminology of our account of character, as long as we grant that many of the subjects in the experiments in question had the right general sort of moral commitments, then the significant phenomenon that the empirical evidence points to is a widespread existence of a gap between one's moral commitments and one's dispositions to act: for a variety of reasons, people often fail to be sufficiently motivated by their moral commitments. The empirical evidence thus does not call into question the existence of character, but simply the existence of strong and good moral characters. This is a problem of moral motivation, and this is a problem that needs to be addressed.

From our previous discussion of the ways in which moral commitments and behavioral dispositions can come apart, we can gain some insight into the likely causes underlying this problem of moral motivation. Where there is a gap between people's moral commitments and dispositions, this is likely because they under-appreciate the demands and pervasiveness of their moral commitments in the face of conflicting demands, or, quite simply, find themselves unmoved by morality—whether this be a result of *akrasia*, *accidie*, or simple selfishness. Most likely it results from a combination of several factors, working at different degrees.

Of all the experiments familiar to the debate over character traits, the most illustrative are the Milgram experiments, for the simple reason that, given the design of the experiments, we have the most information about the subjects' deliberative processes. The details of this experiment are familiar and straightforward: subjects in the experiment were told to ask a person a series of questions and administer shocks to the "learner" if he or she answered incorrectly. The subjects believed they were in engaged in a study designed to test memory and learning, but in actuality, the learners never received actual shocks; the true purpose of the experiment was to test the extent to which people will follow authority. The results of the studies—which were conducted with many different variables across a diverse segment of an international population—showed that the overwhelming majority of subjects followed

authority to the extent that they believed they were administering shocks past the point of "severe danger" to people whose only fault was that they failed to answer trivial questions correctly.

What led subjects in this experiment to engage in such immoral behavior? Many people are tempted to think that this is a case where the subjects believed they were not, in fact, violating any moral standards, and that the morally right thing to do was to follow the orders of the experimenter. However, the lead researcher, Stanley Milgram, found that this simply was not the case. He identified several different factors that enabled the subjects to continue shocking the learner while being fully aware of the harm (supposedly) inflicted by doing so; each of these factors involved some form of self-deception that enabled the subjects to reconcile their immoral behavior with their belief in the immorality of that behavior. The first set of factors consists in the basic strategies of avoidance, where "the subject screens himself from the sensory consequences of his actions," and denial, where the subject rejects "apparent evidence in order to arrive at a more consoling interpretation of events."[33] In these cases, the subject is able to justify her immoral actions only by engaging in a sort of self-deception that lessens the immorality of her actions. A second set of factors involves the transference of responsibility. When subjects were able to place responsibility for their immoral actions either on the learner, or the experimenter, they found themselves able to continue shocking the learner (often with ease), despite the fact that they maintained the belief that it was wrong to do so.[34] A final enabling factor Milgram identified was the process of dissent: subjects who voiced official expressions of disagreement paradoxically found themselves more easily able to continue to engage in their immoral actions.[35]

These factors enabled subjects in the Milgram experiment to engage in behavior that they believed to be immoral; they are all symptoms *not* of the fact that subjects changed or revised their beliefs that it is wrong to inflict harm on others, but rather of the fact that the subjects failed to be properly and sufficiently motivated by these moral commitments. It was relatively easy for these subjects to relegate the demands of their moral commitments to a lower level and to let their behavior be determined by other factors—such as the particular details of the situation. And this, I think, is the lesson to be learned from most of the experiments raised in this debate: situational factors tend to influence our behavior to a surprising degree, the

---

[33]   Milgram 1974, p. 158.
[34]   Milgram 1974, p. 159-161.
[35]   Milgram 1974, p. 161.

result being a significant gap between one's moral commitments and behavioral dispositions.

Given my proposed account of character, and some basic assumptions about the content of people's moral beliefs (namely, that most people have the right, general sort of moral commitments), we can thus conclude that the empirical evidence that Doris and Harman take to show that people lack character really shows that *knowing what is the right thing to do is simply not enough to guarantee right action*. There is a frequent gap between people's moral commitments and behavioral dispositions. The implications of this conclusion for moral theories is clear: moral theorists need to focus not only on formulating and instilling the "right" moral commitments in people, but they also need to focus on closing this gap between one's moral commitments and dispositions. This, I think is, or should be, the distinctive task of character-based moral theories.

§4. Closing the Gap

Social psychologists would describe this task of "closing the gap" between an agent's moral commitments and her dispositions as a task of self-regulation, the process whereby an agent controls her behavior on the basis of her beliefs. In devising strategies for closing the gaps between beliefs and dispositions, then, we can benefit from psychologists' explorations of the most efficient means towards self-regulation. While we cannot here explore all of the details involved in self-regulation, I would like to highlight what psychologists take to be the central obstacles towards self-regulation and offer a brief look at how a moral theory might contain resources to help agents self-regulate and close the gap between their moral commitments and behavioral dispositions.

Karoly argues that the two main problems facing self-regulation are "intergoal conflict and goal imprecision."[36] These problems preclude people from successfully acting on their beliefs, even when they *want* to. As Karoly notes, western society is characterized by the "powerful and derailing or dysregulatory factor" of the "inherent vagueness or fuzziness of life tasks."[36] Nowhere else is this "fuzziness" clearer than in the case of morality. While people may know it is important to treat others well, often times it is unclear just how important it is (Is it more important than pursuing their own goals? Why?), and what exactly "treating others well" requires (Is it more important to fulfill an obligation of timeliness, or to stop to help others?).

---

[36]   Karoly 1998, p. 745.

There is significant consensus amongst social psychologists that the first step in self-regulation is to clear up these confusions. Doing so enables people to go from a position of wanting to act well, a position of "motivational wanting", to a position in which they have the plans essential to success, a position of "volitional trying".[37] When people have detailed plans and strategies for action they are significantly more successful in attaining their goals than when they simply have the end-goal in mind. This claim has been supported by a wide range of studies, ranging from studies on opiate addicts attempting to fulfill a goal of writing a resume, to individuals trying to fulfill New Year's resolutions, to college students trying to get themselves in for a medical check-up.[38]

What can moral philosophers learn from this research on self-regulation? The message, I hope, is clear. First and foremost we need to devise a clear moral theory that avoids the problems of goal conflict and imprecision and offers individuals whom have the goal of moral action in mind strategies for attaining this goal. While now is not the place to explore the details of such a moral theory, let me gesture at what I take to be the most promising approach: this is to develop an account of morality from which agents can derive a clear understanding of the role morality plays for agents, the goals it seeks to accomplish, and how these relate to the moral agent.

When an agent sees that morality is important *to her*, and that *she* should do the right thing, she should be able to form the requisite motivation to do the right thing. This understanding of morality is important for the development of moral commitments and practical judgments. It is also important insofar as it enables the agent to act on her practical judgments, for an agent can use this understanding of morality (a) as her reminder of why it is imperative to do the right thing, even in trying situations that test her resolutions—thus avoiding goal conflict; and (b) to derive a set of strategic rules for action, and a rough decision procedure that enables her to decide amongst tough cases—thus avoiding goal imprecision. Ideally, the agent will be able to develop a belief system that works in unison with and supports her moral commitments and that can enable her to decide what the morally right act is, especially in the face of trying situations where the temptations of self-deception are hard to resist.

To see how this might work, let us suppose that enabling pro-social behavior turns out to be the primary role and goal of morality

---

[37] Karoly 1998, p. 743.

[38] The first two studies are discussed in Brandstatter et al., 2001, the last in Gollwitzer and Oettingen 2000.

and that we can connect the importance of pro-social behavior with an agent's psychological well-being, thus enabling the unified agent to develop resolutions to act in favor of pro-social behavior. We can then use this goal to develop general guidelines for action, such as avoiding harming others, keeping contracts, being honest, etc. These guidelines serve as rules of thumb in ordinary situations, so that on most days, the agent will be able to successfully act on her resolutions by following these rules. In more trying cases, where the rules conflict, or there do not seem to be any applicable rules, the agent can draw on her understanding of the role and goals of morality, and on this basis try to figure out and act on what she thinks is right. One important thing to note about the nature of these rules is that they operate as strategic guidelines, rather than as constraints on behavior. Agents have resolved to do the right thing, and the rules help them figure out what to do in particular situations. They help agents develop strategies for right action, as opposed to telling them simply what they can and cannot do.

Perhaps more important than having a set of general guidelines is that the agent be able to understand, appreciate, and identify with morality on a deep, psychological level, so that she forms the motivation to do the right thing and sees morality as important for her. This psychological identification of morality is what might have helped prevent subjects in the Milgram experiment from continuing to engage in immoral behavior. If they had a clear and present appreciation of the importance of treating people well, and importantly, understood why it was personally important for *them* to treat people well, then it is less likely that they would have engaged in acts of self-deception, or thought that immoral actions were acceptable as long as they were not held responsible for it, or thought that it was sufficient for them to voice dissent, instead of stopping the experiment.

If we are to enable people to develop strong, good moral characters and successfully exhibit them, then moral theories must be accessible to individuals across many dimensions. The central components of such a theory are two-fold. First, it must be connected to an individual's psychology in such a way that the individual can understand, appreciate, and identify with the morality and its requirements; and second, it must provide the agent with the resources to decide what to do in any given situation. I have argued that knowledge of both the overall goal and objective of morality, and a general set of strategic guidelines, will suffice. It is this multi-faceted understanding of morality that helps people to bridge the gap between their moral commitments and dispositions, by providing them with the resources to develop the correct practical attitudes and successfully act on them.

§5. Conclusion

The empirical evidence Doris and Harman bring to our attention is of important consequence for moral philosophy, although not for the reasons they suggest. It shows that most people often fail to do the right thing, and as such it shows that moral theorists need to do more than they have done. We need to focus on the practical question of how we can solve the problem of moral motivation, and work to resolve moral fallibility. Working from an account of character that I think best fits everyday attributions of characters, I have suggested one way of solving this problem: to make it the case that people not only have the "right" sort of moral beliefs, but also have enough knowledge and resources to figure out how to apply them in difficult situations, and have resolved to do the right thing. Most importantly, though, I hope to have put Doris's and Harman's objection into perspective: while we may not agree with their conclusions, there is an important phenomenon here that deserves attention.

References

Abraham, C., Sheeran, P., and Johnston, M. (1998). "From Health Beliefs to Self-Regulation: Theoretical Advances in the Psychology of Action Control." *Psychology and Health*, 13(4): 569–591.

Athanassoulis, N. (2001). "A response to Harman: Virtue ethics and character traits." *Proceedings of the Aristotelian Society*, 100: 215–221.

Brandstatter, V., Lengfeld, A., and Gollwitzer (2001). "Implementation Intentions and Efficient Action Initiation." *Journal of Personality and Social Psychology*, 81(5): 946–960.

Bratman, M. (1979). "Practical Reasoning and Weakness of the Will." *Noûs*, 13: 153–171.

Darley and Batson (1973). "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior". *Journal of Personality and Social Psychology,* 27: 100–108.

Doris, J. M. (1998). "Persons, situations, and virtue ethics." *Noûs*, 32: 504–530.

—— (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.

Gollwitzer, P. and Oettingen, G (2000). "The Emergence and Implementation of Health Goals." In *Understanding and Changing Health Behavior*. Ed. Paul Norman, Charles Abraham and Mark Connor. Amsterdam: Harwood Academic Publishers.

Harman, G. (1999). "Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error." *Proceedings of the Aristotelian Society,* 99: 315–331.

—— (2000). "The nonexistence of character traits." *Proceedings of the Aristotelian Society*, 100: 223–226.

—— (2003). "No character or personality." *Business Ethics Quarterly,* 13: 87–94.

Hume, D. *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton. New York: Oxford University Press, 2002.

Kamtekar, R. (2004). "Situationism and Virtue Ethics on the content of our character." *Ethics*, 114: p. 458–491.

Karoly, P. (1998). "Expanding the Conceptual Range of Health Self-Regulation Research: A Commentary." *Psychology and Health*, 13(4): 741–746.

Kupperman, J. J. (1991). *Character*. New York: Oxford University Press.

—— (2001). "The indispensability of character." *Philosophy*, 76: 239–250.

Martin, M. (1986). *Self-Deception and Morality*. Kansas: University Press of Kansas.

Merritt, M. (2000). "Virtue ethics and situationist personality psychology." *Ethical Theory and Moral Practice*, 3: 365–383.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Miller, C. (2003). "Social psychology and virtue ethics." *The Journal of Ethics,* 7: 365–392.

Regan, T. (2001). "Work, Hypocrisy, and Integrity." *Defending Animal Rights*. Chicago: University of Illinois Press.

Smith, M. (1994). *The Moral Problem*. Cambridge, Mass: Blackwell Publishers.

Solomon, R. C. (2003). "Victims of circumstances? A defense of virtue ethics in business." *Business Ethics Quarterly,* 13: 43–62.

Sreenivasan, G. (2002). "Errors about errors: Virtue theory and trait attribution." *Mind*, 111: 47–68.

Swanton, C. (2003). *Virtue Ethics: A Pluralistic View*. New York: Oxford University Press.

Vranas, P. (2005). "The Indeterminacy Paradox: Character Evaluations and Human Psychology." *Noûs*, 39: 1–42.