ORIGINAL PAPER

The Bleak Implications of Moral Psychology

Edouard Machery

Abstract In this article, I focus on two claims made by Appiah in Experiments in Ethics: Doris's and Harman's criticism of virtue ethics fails, and moral psychology can be used to identify erroneous moral intuitions. I argue that both claims are erroneous.

In Experiments in Ethics, Kwame Anthony Appiah examines the threats and promises that moral psychology (viz. the empirical study of our moral judgments and behaviors) carries for moral philosophy.[1] Experiments in Ethics is a tour-de-force. Written in an elegant and engaging manner, it synthesizes a large amount of empirical knowledge about morality without sacrificing the acumen of the argumentation.[2]

---

[1] In this article, I follow Appiah in distinguishing moral psychology from moral philosophy, but I should note that in my mind, moral psychology is both empirical and philosophical (for a similar view, see the introduction of the forthcoming Handbook of Moral Psychology).
[2] I highly recommend reading the endnotes where much of the research on which Appiah relies is described and referenced.

E. Machery (✉)
Department of History and Philosophy of Science,
University of Pittsburgh,
1017 CL,
Pittsburgh, PA 15260, USA
e-mail: machery@pitt.edu

Moral philosophers and moral psychologists sometimes suggest that the development of the empirical study of morality threatens the philosophical study of moral and ethical matters, although the exact nature of the threat is often left unspecified (but see, e.g., [1, 2]).[3] Appiah's central task in Experiments in Ethics is to exorcise this threat: Not only have moral philosophers nothing to fear from moral psychology; in fact, they need moral psychology to fulfill many of their goals. They need empirical knowledge about the causes of human behavior to determine what kind of person we can hope to become and how this can be done (Chapter 2); they need empirical knowledge about human moral judgments (or, as they are often called, moral intuitions) to distinguish those judgments that provide genuine moral reasons from those judgments that should be disregarded (Chapter 3); and they need empirical knowledge to understand the pluralism of our commonsense moral outlook (Chapter 4).

As an experimental philosopher and as a naturalist philosopher, I rejoice having Appiah as a fellow traveler. I also find his message attractive since it might abate some of the hostility that naturalism and experimental philosophy regularly encounter among philosophers. Unfortunately, I doubt that Appiah's irenic position can be sustained. In this article, I will argue that

---

[3] Appiah distinguishes moral and ethical questions as follows (37): Moral questions concern what we should do; ethical questions concern what life we should live.

Appiah has failed to come to grips with the bleak implications of moral psychology for moral philosophy.

Here is how I will proceed. In "Should we Care About what Person to Be?," I will argue that Appiah's reply to Harman's and Doris's war on virtue ethics does not exorcise the darkest threat posed by the situationist literature. Situationism casts doubts on whether moral philosophers have been justifiably preoccupied with the question of what kind of person we should strive to be. In "A Motley of Dubious Moral Intuitions", I will argue that Appiah is unduly optimist about the prospects of using the empirical literature on morality to distinguish spurious from genuine commonsense moral intuitions.

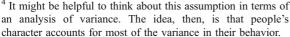**Should we Care About what Person to Be?**

In this first section, I focus on an important goal of moral philosophy: finding out what kind of person we should be. I argue that various bodies of research in psychology undermine this goal.
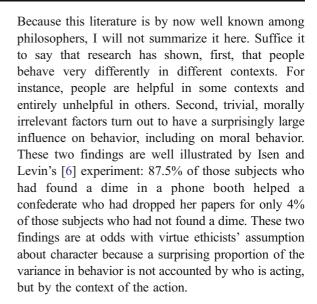
The Situationist Threat

The second chapter of *Experiments in Ethics*, "The Case against Character," focuses on Gil Harman's [3] and John Doris's [4, 5] war on virtue ethics. As I understand them, Harman and Doris argue that many (if not all) virtue ethicists assume that a stable psychological source, our character, underlies much of our behavior.[4] Naturally, virtue ethicists do not claim that people's character is always virtuous. They believe that some people are mean and that they tend to act meanly, that other people are greedy and that they tend to act greedily, and that yet other people have little courage without being through and through wimps and that they tend to act with little courage. In addition, virtue ethicists are aware that contextual factors also influence behavior, but they hold that the causal influence of these factors is not particularly strong. When one looks at someone's behavior, one can see a pattern (due to her character) through the noise due to contextual factors.

Harman and Doris argue that the situationist literature in social psychology belies this assumption.

Because this literature is by now well known among philosophers, I will not summarize it here. Suffice it to say that research has shown, first, that people behave very differently in different contexts. For instance, people are helpful in some contexts and entirely unhelpful in others. Second, trivial, morally irrelevant factors turn out to have a surprisingly large influence on behavior, including on moral behavior. These two findings are well illustrated by Isen and Levin's [6] experiment: 87.5% of those subjects who had found a dime in a phone booth helped a confederate who had dropped her papers for only 4% of those subjects who had not found a dime. These two findings are at odds with virtue ethicists' assumption about character because a surprising proportion of the variance in behavior is not accounted by who is acting, but by the context of the action.

Appiah's Exorcism of the Situationist Threat

In reply, Appiah contends that the situationist findings reviewed by Harman and Doris fail to threaten the most promising version of virtue ethics—virtues are ideals that should regulate our ethical choices ([7], 48–49):

> Philosophical accounts of the character ideal of compassion, the conception of it as a virtue, need make no special assumption about how easy or widespread his deep disposition is. Acquiring virtue, Aristotle already knew, is hard; it is something that takes many years, and most people don't make it. (...) But difficult is not the same as impossible; and perhaps we can ascend the gradient of these virtues through aspiring to the full-fledged ideal.

Since ideals are norms, the fact that people do not reliably act with virtue (courage, generosity, compassion, etc.) and thus do not have virtuous characters is no argument against virtue ethics.[5] What would count as an argument against it would be either to show that it is not the case that virtues *should* be pursued (a normative claim) or to argue that it is *impossible* to become more virtuous (an empirical claim that is not supported by the available evidence). Appiah seems confident that none of these arguments is forthcoming. Moral philosophers have thus

---

[4] It might be helpful to think about this assumption in terms of an analysis of variance. The idea, then, is that people's character accounts for most of the variance in their behavior.

[5] Doris [5] discusses this reply, noting that it raises a specific empirical question: To which extent are people able to comply with these norms?

been rightly preoccupied with understanding what kind of person we should strive to be.

Appiah goes further than this somewhat standard reply on behalf of virtue ethics. He acknowledges that treating virtues as ideals raise various questions concerning how "we human beings [might] take seriously an ideal that human beings must fall so far short of attaining" (50). Particularly, he examines whether a virtue ethicist could recommend adopting various behavioral heuristics. This discussion is fascinating, but, for the sake of space, I will not examine it in this article.

The Situationist Threat: The Return

I now argue that the situationist threat survives Appiah's exorcism. The reason is that Appiah fails to come to grips with the most pressing threat posed by the situationist literature. This threat *cannot* be that very few people have a virtuous character, as Appiah seems to believe it is, because we didn't need situationist psychologists to tell us this. We admire Gandhi and other paragons of virtue precisely because we are already convinced that virtue is rare. So, what is really threatening in the situationist literature? This: The psychological causes of human behavior might not have the proper organization for the central question of virtue ethics, namely "What kind of person to be?," to be an appropriate way of thinking about the norms of human behavior. And *this* threat is not met by arguing that virtues are ideals to be emulated.[6] I now elaborate on these points.

The question, "What kind of person to be?," is central to virtue ethicists' approach to moral and ethical matters (although, as Appiah convincingly argues, every major moral philosopher has thought about it). Virtue ethics proposes to shift emphasis from the issue of how one should act to the issue of what type of person to be (some virtue ethicists even propose to reduce the former issue to the latter).

One might object that for some virtue ethicists (e.g., [9]), the goal of virtue ethics is not to determine what kind of person we should be. Rather, these virtue ethicists are primarily concerned with how one should act: In substance, in any context, it is right to

do what the virtuous agent would do. There are two mutually consistent replies to this objection. First, I do not mean to challenge any conceivable version of virtue ethics. (Nor did Harman and Doris.) I mean to challenge those influential versions of virtue ethics that are concerned with determining what kind of person we should be. Second, the virtue ethicists alluded to in the objection seem to forego the most original insight in virtue ethics: Agents should work on changing the kind of person they are. Appiah develops a similar criticism of these virtue ethicists in Chapter 2 of *Experiments in Ethics*, and I agree with his critical assessment.

But what does it mean to have a particular character or be a particular kind of person? Having a given character or being a particular kind of person just is to have some specific values, norms, (first- and second-order) desires, beliefs, moods, emotions, etc. It is not something one might have in addition to these values, norms, etc. Characters and kinds of person are meant to differ from one another in that people who have different characters or who are different kinds of person have different values, norms, desires, emotions, etc. A good person has different values, norms, desires, etc., than a bad person.

Furthermore, asserting that there are characters and kinds of person assumes that the psychological causes that are meant to constitute our character and the kind of person we are (our values, desires, norms, emotions, etc.) have a specific causal structure: They (or at least many of them) are *unified*. That is, they are causally influenced by a common cause or they causally influence one another. As a result, the values of any one of them depend on the values of the other causes and are correlated with them.[7] What the good person desires is related to what emotions she is disposed to have, what moods she typically has, the norms she harbors, etc.[8]

---

[6] It does not matter whether Harman [3] and Doris [4, 5] originally conceived of the situationist threat in these terms. Doris's [8] arguments are close to the argument presented here.

[7] It might be helpful to think about this in terms of factor analysis. Suppose people really have characters. Then, if one were to analyze someone's behavior factorially, a single factor would emerge (in contrast to several factors corresponding to the diverse psychological causes).

[8] To clarify, the issue does not concern the unity of the virtues (viz. whether courage, generosity, temperance, etc., necessarily go together). Rather, it concerns the kind of unity among beliefs, desires, values, emotions, etc., that is required for someone to have a given character trait. For instance, to be courageous is to be disposed to have specific beliefs, desires, values, etc., with respect to situations of danger.

Why is that so? Remember that the notions of character and of kind of person are meant to explain why behaviors differ (because characters differ) and how to change people's behavior (change their character). But now suppose that the mental states and dispositions that constitute our character and the kind of person we are are not unified. Then, one would not explain why behaviors differ by referring to people's character; rather, one would refer to their emotions, or to their values, or to their moods—viz. to specific psychological causes. Similarly, one would not propose to change people's behavior by changing their character; rather, one would propose to intervene on their moods, emotions, values, second-order desires, and so on. It would then seem that people have no character.

As I read it, the situationist literature suggests that, *to some extent*, the psychological causes of human behavior are not unified. If they causally influenced one another or if they were causally influenced by a common cause, then changes in one of them would involve changes in the other ones. But the situationist literature suggests that the values of at least some psychological causes (e.g., moods) vary independently from the values of the other ones. Consider again Isen and Levin's dime experiment. Those participants who found a dime probably helped because their mood was improved by their finding. However, it is unlikely that their beliefs (including their normative beliefs), their desires, their values, etc., were changed. Thus, mood, which has a potent influence on behavior, seems to vary independently of the other mental states and dispositions that are meant to constitute our character and the kind of person we are. I will say that agency is *partly disunified* when some psychological causes of human behavior are allowed to vary freely.[9]

If these causes were *largely* disunified, then there would be no such thing as a kind of person. It would then be at least unclear whether there are norms about what kind of person to be and virtue ethicists' normative interest in what kind of person to be—an interest shared by Appiah—would be an inappropriate way of approaching the norms of human behavior. Clearly, the situationist findings do not show that agency is disunified to this extent, but they suggest that this is a live possibility. *That* is the darkest threat raised by situationism. In addition, the claim that virtues are

ideals fails to address this threat. If there is no such thing as a kind of person, then it is unclear whether the recommendation that people be more honest, compassionate, and courageous is meaningful, since it asks people to become a different kind of person.

Appiah could reply in at least two different ways. First, conceding for the sake of the argument that agency is largely disunified, he might reply that a normative interest in what kind of person to be remains meaningful if the psychological causes of human behavior can be unified to a greater extent than they currently are. And, he would add, there is no evidence that this is impossible. This reply has two main shortcomings. First, a mere logical possibility is insufficient: We want some evidence that these psychological causes can be unified to a point where it makes sense to refer to people's character. Second, the fact that some psychological causes seem beyond our control (see below) does not bode well for the proposal that their unity can be strengthened. If we cannot intervene on the mental states and dispositions that are meant to constitute our character or the kind of person we are, then how can we bring them in step with one another?

Alternatively, Appiah might reply that the empirical literature suggests that the agency is to a large extent unified and that its unity can be further strengthened. He could find ammunitions in various fields of psychology. Personality psychologists assume that behavior has a certain causal unity: The personality dimensions they identify are meant to account for a substantial proportion of the variance in people's behavior (e.g., [10]). If one focuses only on the situationist literature, it is also unclear whether situationists have really undermined the notions of personality and character [11, 12]. There is finally a long tradition in educational psychology that attempts to identify the practices furthering a good character (e.g., [13, 14]).

This is certainly not the place to examine these literatures. Rather, I want to do two things. In the next section, I will review some further empirical findings that support the claims that the psychological states and dispositions that are meant to constitute character and the kind of person we are are partly disunified and that it might be difficult to bring them in step with one another. Before, however, I would like to note that by endorsing the second reply, Appiah would acknowledge that the consilience of moral psychology and moral philoso-

---

[9] Of course, agency can be more or less unified.

phy might not be possible: If the empirical literature were to reveal that human agency is disunified, then a large part of moral philosophy would be threatened. In addition, he would acknowledge that that the proper response to the situationist threat involves examining the empirical literature on agency in detail. There is no easy way for moral philosophers out of a laborious study of human behavior.

More Evidence for a Disunified Agency

Much of the recent research in psychology suggests that behavior is the product of numerous causes that are not correlated with one another. Some of these causes may also evade control, either because we are ignorant of them or because we do not know how to influence them. I will briefly consider two bodies of research: the influence of automatic processes on social behavior and the research on implicit biases.

Much research in contemporary social psychology (including many so-called dual-process theories) concur in hypothesizing that our judgments and actions are often driven by unconscious and automatic systems (e.g., [15–17]). Bargh and Williams write ([17], 1):

Much of social life is experienced through mental processes that are not intended and about which one is fairly oblivious. These processes are automatically triggered by features of the immediate social environment, such as the group memberships of other people, the qualities of their behavior, and features of social situations.

These systems (sometimes referred to as System 1) stand in contrast with those systems whose activation is under our intentional control (sometimes referred to as System 2). Importantly for our present purposes, automatic and controlled systems work independently of each other. We are thus endowed with several distinct systems that are independent from each other and that compete to control our behavior.

Appiah is well aware of this literature, but he does not bring it to bear on the debate about character. However, this literature suggests two relevant things. First, some psychological causes—viz. the automatic and controlled systems—are not unified: They are not causally dependent on some unified cause and they do not causally influence one another. Second, we have no direct control over some psychological causes—

namely over the automatic systems—suggesting that it might be difficult to bring them in step with the other states and dispositions that are meant to constitute character.

One could object that there is plenty of evidence that the automatic systems can be controlled (e.g., [18, 19]). There are two mutually consistent replies to this objection. First, the extent to which we can control them is unclear. For instance, control seems very difficult when we are tired or when we have to decide very quickly. In addition, because control is effortful and might deplete our mental resources, control might be often followed by a lack of control. Second, and more important, controlling the automatic processes is one thing, changing them is another one. It is unclear whether merely controlling the expression of some of the states and dispositions that are meant to constitute character and the kind of person we are (rather than changing them) counts as changing the kind of person we are and thus as meeting the virtue ethicist's requirement.

Second, research on stereotyping (particularly, on racism) has revealed that people often have implicit biases—biases they are not aware of having—in addition to the explicit biases they might harbor (for review, see [20]; for discussion in a philosophical context, see [21–23]). Strikingly, people without explicit biases might exhibit some implicit biases, even when they sincerely attempt to be unbiased. Research on the so-called weapon bias illustrates the influence of these biases on behavior. White and African–American subjects are more likely to misidentify a harmless object as a gun if they are first shown a picture of an African–American, rather than a picture of a White-American (for review, see [24]). Payne [25] has shown that the weapon bias is stronger in implicitly biased people. The discovery of implicit biases and of their influence on behavior provides more evidence that the psychological causes of human behavior are disunified: We find at least two sets of opposing racial attitudes within a single agent. The fragmentation is in fact more extensive, because there are different types of implicit attitudes [23]. In addition, research on the malleability of implicit attitudes suggests that at least some attitudes are easy to acquire and difficult to eliminate [26]. Furthermore, various contextual factors (stress, fatigue, alcohol, need for a quick reaction) modulate our capacity to control the expression of these implicit

biases. This tentatively suggests that it might be difficult to bring implicit biases in step with our explicit attitudes.

It is currently unclear whether agency is so disunified that it does not make sense to speak of people's character, but in light of the evidence just alluded to, this is clearly a live possibility. This threatens the normative interest in what kind of person to be. It is important not to misinterpret this claim, however. This argument is not meant to show, and does not show, that the philosophical interest in how one should act is misguided. One can learn to act generously (or without biases) even if there is no such thing as a generous (or unbiased) character. What it shows is that the answer to the question, "How should we act?," might not involve changing what kind of person one is.

A Motley of Dubious Moral Intuitions

Much of experimental philosophy and of moral psychology focuses, not on behavior, but on moral judgments. In Chapter 3, "The Case against Intuition," Appiah examines whether this research threatens moral philosophy. He not only argues that it does not, but he also proposes that this research can be used to identify the intuitions that are genuine reasons for action and that provide the basis for moral theorizing (for a related argument, see [27]). In this section, I challenge this claim.

Appiah's Wide Reflective Equilibrium

Moral intuitions (viz. judgments about specific cases) play an important role in moral philosophy.[10] Moral and ethical theories are bound to agree with many commonsense moral intuitions. As Appiah notes (111–112), people would probably have difficulty assenting to a theory at odds with many commonsense moral judgments. At the same time, however,

no moral or ethical theory can be consistent with all our intuitions: Some intuitions have to be jettisoned.[11] At this point of the argument, Appiah argues that the most commonly endorsed method for justifying moral and ethical theories among moral philosophers—reflective equilibrium—is ill-suited for choosing between the moral intuitions that should be jettisoned and those that should be preserved.

By contrast, Appiah proposes that empirical studies can fulfill this sorting function (110–111; see also 115):

> Understanding where our intuitions come from can surely help us to think about which ones we should trust. (...) The proposal that—to put it very crudely—it's our feelings that guide us to the intuition about the footbridge case, while our reason guides us in the original trolley problem is the right sort of thing we might want to consider in deciding whether that intuition is right.

Appiah does not explain in much detail how the empirical study of our moral intuitions can help us determine which intuitions are spurious. Rather, he reviews several studies on moral judgments in order, I assume, to illustrate how this can be done.

He first considers the research on framing effects, focusing on Tversky and Kahneman's [28] famous Asian flu case. People are presented with four possible ways to prevent the outbreak of a deadly disease. Although these four options have the same expected utility, people typically do not treat them as being equivalent. Furthermore, these options form two pairs of identical choices (only the wording differs within a pair). Because people distinguish identical choices, one can clearly assert that in this case, intuitions are irrational.

Appiah also reviews some research on the cognitive mechanisms and emotions that underlie moral judgments, illustrating this research with Joshua Greene's well-known work on the trolley cases [29–31]. He proposes that knowing the nature of the mechanisms that deliver our moral intuitions (emotions, unconscious, automatic process, conscious, deliberative reasoning, etc.) can enable philosophers to evaluate

---

[10] Interestingly, in the last chapter of *Experiments in Ethics*, Appiah argues that the importance of intuitions in moral philosophy is regrettable. For simplicity, I will overlook this claim in what follows.

[11] As Appiah remarks, moral theories have to walk a tight rope between being too conservative (agreeing with many commonsense moral intuitions) and being too reformative (disagreeing with many moral intuitions).

whether these intuitions are genuine or whether they should be jettisoned.

It is noteworthy that in spite of Appiah's pointed critique of reflective equilibrium, he ultimately recommends a particular version of it. It is common to distinguish narrow and wide reflective equilibrium. In at least one interpretation, what distinguishes the latter from the former is the use of empirical knowledge in addition to our intuitions about cases and our pre-theoretical principles. Appiah is naturally understood as inviting philosophers to include empirical knowledge about the cognitive mechanisms that produce moral intuitions among the beliefs considered in wide reflective equilibrium.

Moral Psychology is no Intuition Sieve

As we just saw, Appiah proposes that the properties of the systems delivering our moral intuitions—whether or not they are automatic, computationally limited, or conscious, whether or not they neglect some cues, whether or not they can learn, and so on—are relevant for deciding whether the delivered intuitions are genuine or spurious. However, it is not obvious how to use these properties to identify spurious intuitions in a non-circular manner. Consider the automaticity of some intuition-delivering systems. Do automatic processes deliver genuine or spurious intuitions? Answering this question on a purely factual basis seems impossible: What is the factual connection between automaticity and moral genuineness?[12] Rather, to answer this question, one must appeal to one's moral norms: Automatic systems deliver genuine moral intuitions if these intuitions are in conformity with our moral norms. But, in the present context, this is plainly circular since our goal is to distinguish genuine from spurious intuitions in order to determine what our moral norms should be. The same circularity is found when one considers the properties of other intuition-delivering systems.

Maybe, Appiah would concede that his proposal involves some kind of circularity, but deny that this circularity is vicious. Briefly, I note first that Appiah should then explain why this circularity is not vicious. Second, this circularity does seem vicious because the circle of mutual support between (say) the proposition

that automatic systems yield spurious intuitions because these intuitions are inconsistent with moral norms and the proposition that some putative moral norms are genuine in spite of being inconsistent with the intuitions delivered by automatic systems is very narrow indeed.

Let me illustrate the argument developed here with a few examples, starting with the commonsense intuitions about happiness. Like most philosophers, Appiah distinguishes happiness (or a fulfilled life) and subjective satisfaction by appealing to our intuition that these two conditions differ, as is shown by Nozick's famous "experience machine" thought experiment. However, recent research in experimental philosophy shows that folk intuitions about happiness form a complicated pattern. Felipe de Brigard [33] asked people to imagine that they discover that, unbeknownst to them, they have spent their life plugged in the experience machine. Then, people were asked whether they would want to stay plugged or rather whether they would want to unplug. Surprisingly, de Brigard found that most people would not unplug from the experience machine, and this even when they were told that they were exchanging their virtual life for a (real) life of leisure! De Brigard's research suggests two things. First, in some cases, subjective happiness trumps reality: In some cases, people prefer an illusory happiness to a subjectively unhappy life. Second, people's judgments in Nozick's original thought experiment and in de Brigard's variants are influenced by the fact that people are conservative: They are reluctant to trade what they have for something else (a phenomenon known as "the status quo bias" in psychology). So, are our intuitions about happiness genuine? Does the finding that conservativeness influences our intuitions about happiness show that these intuitions are spurious? It is unclear how to answer these questions because there is no clear factual connection between conservativeness and moral genuineness and because one would reason circularly if one were to note that moral intuitions that are influenced by conservativeness violate some moral norms.

Consider also the research done by Balázs Gyenis and myself [34]. We gave people a variant of the standard trolley case, in which the character does *not* push the switch, and we asked them to evaluate how appropriate the character's decision was. We varied the ratio of, and the difference between, the number of people killed if the trolley is allowed to continue on the main tracks and the number of people killed if the trolley

---

[12] Greene seems to hold that automaticity goes with irrationality, while others propose that it goes with rationality (e.g., [32]).

is sent to the sidetracks. Our goal was to examine whether people's judgments were sensitive to different ratios and differences. To our own surprise, we found that people gave roughly the same answer for widely different ratios and differences. Now, let me ask again: Are our intuitions about causing harm to prevent a greater harm rendered spurious by the fact the mechanisms that deliver them are (at least sometimes) insensitive to the quantitative trade-off between the harm caused and the harm avoided?

Or consider the fact that recent research in experimental philosophy and in psychology suggests that the intuitions triggered by the trolley cases are influenced by cues whose moral status is unclear. Alison Stuart and Eric Hatleback [35] have recently argued that our intuitions about the moral permissibility of causing some harm in order to prevent some greater harm is influenced by our felt similarity with the individuals to whom harm is caused. People are less likely to judge that it is permissible to cause some harm in order to prevent a greater harm when they feel that they are similar to the individual harmed. In recent work, Greene and colleagues [36] have also shown that intuitions about moral permissibility are influenced by whether the action involves using one's physical strength to cause harm. People are more likely to judge that an action that prevents a greater harm by causing a smaller harm is impermissible if one uses one's physical strength to cause harm. Now, again, are our intuitions rendered spurious by the fact the mechanisms that deliver them are influenced by this kind of cues?

It is instructive to compare these patterns of intuitions to the pattern of intuitions triggered by the Asian flu case. In the latter case, the unassailable conclusion that people's intuitions are spurious does not rest on any finding about the properties of the mechanisms delivering these intuitions; rather, it appeals to the fact people distinguish between equivalent cases. By contrast, Appiah proposes to rely on information about the mechanisms delivering our intuitions (e.g., whether they are influenced by conservativeness…) to determine whether these intuitions (e.g., our intuitions about happiness or our intuitions about causing harm to prevent a greater harm) are genuine. But it is dubious whether this proposal can be implemented in a non-circular manner.

Note finally that our incapacity to sort our moral intuitions leaves moral philosophers and, more generally, agents in a troubling situation. From a phenomeno-logical point of view, moral intuitions have some kind of clout: They seem to provide us with a reason to act one way rather than another. However, two facts raise some doubts about whether they do provide such reasons: We know that some, and maybe many, moral intuitions are spurious (as is shown by the Asian flu case), but, save a few cases, we do not know how to identify the intuitions that ought to be jettisoned.[13]

Conclusion

Experiments in Ethics is an important book. It may convince traditional moral philosophers that they need to get acquainted with the empirical study of morality. That would be quite a feat! And, as we have come to expect from Appiah, it is also beautifully written. Still, Appiah's irenic project faces serious, perhaps insurmountable, obstacles. Various empirical findings suggest that agency is to a surprising extent disunified. If agency were truly disunified and if we were unable to strengthen its unity, speculating about what kind of person people to be—an important tradition in moral philosophy—would be an inappropriate way of thinking about the norms of human behavior. Furthermore, psychology has shown that at least some moral judgments are spurious, but, except in a few cases, it has not given us the means to distinguish genuine from spurious judgments.

References

1. Levy, N. 2006. Cognitive scientific challenges to morality. *Philosophical Psychology* 9: 567–587.
2. Sinnott-Armstrong, W. 2008. Framing moral intuitions. In

---

[13] Examining the implications of this situation is beyond the scope of this article.

Moral psychology, volume 2: The cognitive science of morality, ed. W. Sinnott-Armstrong, 47–76. Cambridge: MIT Press.

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

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

5. Doris, J.M. 2002. *Lack of character*. Cambridge: Cambridge University Press.

6. Isen, A.M., and P.F. Levin. 1972. The effect of feeling good on helping: cookies and kindness. *Journal of Personality and Social Psychology* 21: 384–388.

7. Appiah, K.A. 2008. *Experiments in ethics*. Cambridge: Harvard University Press.

8. Doris, J.M. Forthcoming. *A natural history of the self*. Oxford: Oxford University Press.

9. Hursthouse, R. 1999. *On virtue ethics*. Oxford: Oxford University Press.

10. McAdams, D.P., and J.L. Pals. 2006. A new big five. Fundamental principles for an integrative science of personality. *American Psychologist* 61: 204–217.

11. Epstein, S. 1979. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology* 37: 1097–1126.

12. Sabini, J., and M. Silver. 2005. Lack of character? Situationism critiqued. *Ethics* 115: 535–562.

13. Dewey, J. 1897. My pedagogic creed. *School Journal* 54: 77–80.

14. Prestwitch, D.L. 2004. Character education in America's school. *School Community Journal* 14: 139–150.

15. Greene, J.D., and J. Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6: 517–523.

16. Wilson, T.D. 2002. *Strangers to ourselves. Discovering the adaptive unconscious*. Cambridge: Harvard University Press.

17. Bargh, J.A., and E.L. Williams. 2006. The automaticity of social life. *Current Directions in Psychological Science* 15: 1–4.

18. Monteith, M.J., J.W. Sherman, and P.G. Devine. 1998. Suppression as a stereotype control strategy. *Personality and Social Psychology Review* 2: 63–82.

19. Richeson, J.A., and J.N. Shelton. 2007. Negotiating interracial interactions: cost, consequences, and possibilities. *Current Directions in Psychological Science* 16: 316–320.

20. Nosek, B.A., A.G. Greenwald, and M.R. Banaji. 2007. The implicit association test at age 7: A methodological and conceptual review. In *Automatic processes in social thinking and behavior*, ed. J.A. Bargh, 265–292. Philadelphia: Psychology Press.

21. Kelly, D., and E. Roedder. 2008. Racial cognition and the ethics of implicit bias. *Philosophy Compass* 3: 522–540.

22. Kelly, D., E. Machery, and R. Mallon. 2010. Racial cognition and normative theory. In *Handbook of moral philosophy*, ed. J.M. Doris et al. Oxford: Oxford University Press.

23. Faucher, L., & Machery, E. (2009). Racism: against Jorge Garcia's moral and psychological monism. *Philosophy of Social Sciences* 39: 41–62.

24. Payne, B.K. 2006. Weapon bias: split-second decisions and unintended stereotyping. *Current Directions in Psychological Science* 15(6): 287–291.

25. Payne, B.K. 2005. Conceptualizing control in social cognition: how executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology* 89: 488–503.

26. Gregg, A.P., B. Seibt, and M.R. Banaji. 2006. Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology* 90: 1–20.

27. Doris, J.M., and S. Stich. 2005. As a matter of fact: Empirical perspectives on ethics. In *The Oxford handbook of contemporary analytic philosophy*, ed. F. Jackson and M. Smith. Oxford: Oxford University Press.

28. Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–458.

29. Greene, J.D., S.A. Morelli, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107(3): 44–54.

30. Greene, J.D., L.E. Nystrom, A.D. Engell, J.M. Darley, and J.D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.

31. Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.

32. Woodward, J., and J. Allman. 2008. Moral intuition: its neural substrates and normative significance. *Journal of Physiology-Paris* 101: 179–202.

33. De Brigard, F. (2010). If you like it, does it matter if it's real? *Philosophical Psychology* 23: 43–57.

34. Gyenis, B., & Machery, E. (2008). Unpublished data.

35. Stuart, A., & Hatleback, E. (ms). Identification, intuition, and moral dilemmas.

36. Greene, J. D., Cohen, J. D., Nystrom, L. E., Lindsell, D., Clarke, A. C., & Lowenberg, K. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111: 364–371.