

## Naturalism's Perils, Naturalism's Promises: A Comment on Appiah's *Experiments in Ethics*

Jonathan M. Weinberg · Ellie Wang

Received: 4 June 2008 / Accepted: 3 November 2008 / Published online: 17 April 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** In his *Experiments in Ethics*, Appiah focuses mostly on the dimension of naturalism as a naturalism of *deprivation* - naturalism's apparent robbing us of aspects of the world that we had held dear. The aim of this paper is to remind him of that naturalism has a dimension of *plenitude* as well - its capacity to enrich our conception of the world as well. With regard to character, we argue that scientific psychology can help provide a conception of character as dynamic, in a way that may preserve many key aspects of eudaimonistic ethics from the situationists' challenge. With regard to intuition, we address Appiah's worry that naturalistic explanations of the sources of our intuitions may leave us feeling that those intuitions have been thereby debunked. We suggest that it may be that feeling of debunking that should itself be debunked.

**Keywords** Intuitions · Naturalism · Virtue ethics · Situationism · Experimental philosophy

How charming is [naturalistic] philosophy  
Not harsh and crabbed, as dull fools suppose,  
But musical as is Apollo's lute,  
And a perpetual feast of nectared sweets,  
Where no crude surfeit reigns.  
—adapted rudely from Milton, *Comus*

### Two Faces of Naturalism

We welcome Prof. Appiah's engagement with the experimental literature, and it is heartening to see a philosopher of his caliber trying to work carefully through the implications of naturalism for philosophical practice. We expect that at least some philosophers will find Appiah to have erred on the side of science, and has invited the empirical camel's nose under the metaphysician's tent. However, our main concern is that he has not naturalized himself *enough*. Appiah primarily takes science to be offering challenges to philosophy, in taking two early chapters to represent *challenges* to character and intuition, respectively. But in a way, this is a failure to properly internalize the lesson of his first chapter: our philosophical modes of inquiry have always—always already—existed in dialogue with our empirical modes. We want to gesture towards a way to see naturalism as presenting not a challenge, or at least not *just* a challenge, but also offering positive promise as well. To do so requires seeing both of two faces that naturalism presents to philosophy.

---

J. M. Weinberg (✉) · E. Wang  
Department of Philosophy and Program in Cognitive Science,  
Indiana University,  
Bloomington, IN, USA  
e-mail: jmweinbe@indiana.edu

E. Wang  
e-mail: huawang@indiana.edu

Appiah works primarily with a naturalism of *deprivation*: the naturalist's job is to pop the metaphysician's balloon, and keep her feet planted back on (empirically) solid grounds. It is a naturalism of denial, of attenuation. Our fellow experimental philosopher Ron Mallon talks about how nonvegetarians often imagine vegetarianism: they picture an ideal steak dinner, with a baked potato and maybe a bit of broccoli on the side, and then just subtract out the steak. The resulting meal does not seem like much of a meal. Similarly, the naturalist's diet is cooked on Hume's bibliophlogous bonfire, with all the fat ruthlessly trimmed away by Ockham's Razor wielded with Quinean ruthlessness, and it can seem just as problematically unappetizing. Some of the literature that Appiah engages with clearly manifest this same form of naturalism, in John Doris' attempted refutation of the entire tradition of virtue ethics, or Machery *et al.*'s attacks on philosophers' uses of intuitions.

But there is also a naturalism of *plenitude*: stocking our ontological pantry with strange and wondrous ingredients never dreamt of in armchair philosophy. The sciences are mostly there to tell us what there *is*—only very secondarily do they tell us what there *isn't*. It offers up the bizarreness that is modern physics, with the audacity to tell us that *that* strange mess is what lies at the bottom of it all. It fills our psychological larders with all manner of heuristics, biases, and modules, all hidden to view. All of this is alien to our ordinary and armchair perspectives, and it all makes fantastic ingredients with which to cook up some philosophy. The line we are arguing in this paper is that Appiah should do a bit more to embrace this latter aspect of naturalism as well, and if he does so, it might provide at least some answers to some of his worries.

### Naturalism's Case for Character

Appiah represents the relationship between empirical psychology and eudaimonistic ethics as one in which the former denies us something we had dearly wanted from the latter—a picture of the virtues as long-standing, durable, humanly-attainable capacities for doing what we ought, because “most people simply didn't display such multi-track, context-independent dispositions at all, let alone in a unified ensemble” (p.

38). Now one might debate whether the empirical literature on the whole really shows that, but even granting Appiah his interpretation of that literature, we are concerned that he has missed out on the ways in which the scientific psychology can give us a *richer* account of the virtues than the one it may have displaced.

For, although the experimental results may tell us that one observes mostly local traits, and robust global traits are not so widely seen, it is not clear that the experimental results really tell us that the best *conception* of character is one on which it is fragmented, simply a tangle of local traits. The problem we see with this conception is that the structure of character in this picture seems overly static, with no room for the activities of the agent either while facing a moral choice at a time, or in shaping their own moral development.

It is uncontroversial that truly virtuous people are rare. However, this fact does not preclude virtue ethicists from talking about people who possess virtue in a *relative* sense, with more-or-less stable virtues. The fact that virtue may be possessed in degrees becomes especially apparent when we acknowledge that ethicists' idea of possession of virtue not only regards performing virtuous actions in situations, but is better explained as demonstrating multi-track dispositions as Hursthouse espoused (cited by Appiah):

[An honest person] chooses, where possible, to work with honest people, to have honest friends, to bring up her children to be honest. She disapproves of, dislikes, deplores dishonesty, is not amused by certain tales of chicanery, despises or pities those who succeed by dishonest means rather than thinking they have been clever, is unsurprised or pleased (as appropriate) when honesty triumphs, is shocked or distressed when those near and dear to her do what is dishonest and so on. (34)

As we can imagine, all these dispositions may come in degrees. It may be difficult sometimes to compare one person to another and see who fares better as a whole, but that is hardly ethicists' main concern anyway. What traditional virtue ethicists need for the their virtuous ideals to be empirically unproblematic is thus that (a) global virtuous character traits are, even if not *typically* attained, then

possibly attained within human psychology, and (b) virtuous character traits may be developed and one can grow relatively more virtuous by treating virtue as a moral ideal.

First consider how (b) is possible. It is conceivable that having desires or beliefs about, or even a strong commitment to, having a certain character trait gradually increases the probability of performing the trait-relevant outer behavior (in a broad sense) and inner activity (including one's cognitive interpretations and emotional responses to situations, patterns of deliberations, deliberatively or unconsciously formed habits, and so on) in trait-relevant eliciting conditions, also depending on how the situation is construed. For example, when a person has a commitment to be honest, she may try to improve herself in multiple ways to form multi-track dispositions. She may still cheat on occasions, knowingly or unknowingly, and possibly under the influence of situational factors. However, if some time after or during cheating she has a chance to realize what she just did, she might reflect. Given her commitment to honesty, she would not make, or would at least try not to make, similar mistakes again. Moreover, she may keep reinforcing her multi-track dispositions to be honest, since she notices that there is still room to improve. One's commitment and beliefs may very well figure in a change in one's behavioral patterns. Therefore, it seems that there are ways one may gradually develop a relatively stable character (trait) due to her certain belief, desire or commitment.<sup>1</sup>

If one thinks of character traits as just dumb dispositions to behaviors, then one may not be able to see how a person's agency can play this role. But ethicists and psychologists may instead consider character as a *dynamic* structure, in two different ways, one synchronic, the other diachronic. First, one's beliefs, desires, commitments, habits, etc. all exert different strengths and have different weights in forming one's behavior, and character is a matter of achieving a (comparatively) stable resultant vector from those many forces. Second, it allows for the shifting of the balance between these forces over time,

<sup>1</sup> It seems to us that having a commitment does not necessarily imply being conscious of that commitment. One can be aiming at something explicitly or simply endorse it in a less 'conscious' form. However this issue does not have to be settled here.

or when no balance had yet been found, it allows for the creation of new stable configurations. (And, indeed, it can make sense of losing such configurations over time as well, as in cases of mental illness or moral corruption.) This dynamic idea of character is consistent with the experimental data, since it makes no predictions at all about the particular causal make-up of any agent at a time. At the same time, it is not committed to the fragmentation view of character that Doris has proposed. The dynamic conception is also compatible with the globalist conception of perfectly virtuous character, since it is possible that the structure of character will in the end become stable enough which issues considerable behavioral consistency. This allows for at least the possibility that (a) requires, though whether one can see human development as one that can admit of that possibility in some sort of limit is a matter for further empirical investigation.

We think that one can see something like this dynamic conception of character in classic virtue theories themselves. Virtue ethicists' conception of character indeed includes the integration of one's belief, desire, and commitments. As Rachana Kamtekar 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 [1].

We can see that virtue ethicists do not see character as a collection of static traits mechanically issuing in behaviors. Rather, their characterological moral psychology is more committed to the idea that character is a dynamic motivational system consisting of different factors, and may stabilize in the end through the work of practical reason.

One can also find something like this conception of character in paradigm situationist theorizing. For example, after discussing the power of situational factors in affecting one's behavior shown in the empirical data, Ross and Nisbett point out that individual psyches must be understood as systems in a state of tension:

The analysis of any given stimulus situation must include the recognition first that "behavior has to

be derived from a totality of coexisting facts,” and second that “these coexisting facts have the character of dynamic field insofar as the state of any part of this field depends on every other part of the field” (Lewin, 1951, p.25). No simple mechanistic laws relating particular stimuli to particular responses are possible, given that both are always embedded in dynamic contexts that alter and constrain their effects [2].

Such a dynamic picture of the psyche does not of itself affirm or deny the existence of enduring resultant behavioral dispositions in the system. It is thus possible that in this tension system there are no stable dispositions such as global personality traits, and any situational perturbation may tilt the system. Thus the idea of a tension system does not automatically confirm the existence of virtue. Also, virtues are typically represented as the product of a harmonious, mutually-reinforcing set of motives, reasons, and habits, whereas a tension system, as the name suggests, may involve different systems in competition with each other. However, the virtue ethicist’s way of thinking about character and the situationist psychologist’s may still share a common ground that is distinct from the fragmentation conception.

Moreover, though the impact of situational factors on behavior may be at times unexpected and powerful, there is still room to find individual differences in this regard, and some form of behavioral consistency for at least some persons. Ross and Nisbett also point out that the better conception of individual differences is argued by many<sup>2</sup> “to be found in the enduring motivational concerns and cognitive schemes that guide attention, interpretations, and the formulation of goals and plans.” One may be consistent in “pursuing consistent goals using consistent strategies, in light of consistent ways of interpreting [her] social world.” (Ross and Nisbett, p.20) This may be the “consistency” we are looking for when we define one’s character. Compare to another passage from Kamtekar:

The virtuous character that virtue ethics hold up as an ideal is one in which these motivations are organized so that they do not conflict, but

support one another. Such an organization would be an achievement of practical reason, and its behavioral manifestation would be cross-situational consistency (in a sense that is somewhat different from the situationists’) [1].

If we interpret the consistency among the goals one pursues, the consistency among strategies one uses, and the consistency in one’s interpretations of her social world to all be the work of practical reason, the social psychologists’ view of consistency surprisingly maps onto the kind of consistency virtue ethicists have in mind.

Here we can again see that the psychologists’ idea of a person’s psyche as a tension system and virtue ethicists’ conception of character as a dynamic motivational structure operate share a deep commitment to a framework of complexly interacting psychological factors, and not just an agglomeration of unthinking behavioral dispositions. Both consider one’s beliefs and goals to be factors of the motivational structure and cannot be overlooked when we consider one’s character. The static fragmentation view of character ignores these aspects and thus does not line up well with either conception. As mentioned, the idea of tension system does not confirm the existence of virtue or that it can be developed, but it directs us how these claims may be tested. Virtue ethicists typically emphasize the importance of practical reason in virtue development and integration. It would be interesting and important to see how we may capture this idea in the framework of a dynamic character structure.

This idea of a tension system, gleaned from the psychological literature, provides compelling resources for wrestling with some elements of the challenge to character that Appiah is troubled by. First, this psychological picture offers a way to situate a theoretical claim that Appiah finds compelling, but seems unable to make a home for, namely, that “the understanding of virtue required by a viable ethics is not the globalist one.... Individual moments of compassion and moments of honesty make our lives better, even if we are not compassionate and honest through and through” (70). On the dynamic conception of character, that a particular configuration of motives, thoughts, and habits is an instance of (say) compassion in no way requires that it be more than fleetingly stable. This is not what traditional virtue ethicists have sought, but it can make sense of applying the eudaimonistic framework even when such enduring traits are not to be found. What

<sup>2</sup> For example, Ross and Nisbett listed George Kelly (1955), Mischel (1973), Markus (1977), Markus, Smith, & Moreland, (1985), and Cantor & Kihlstrom (1987).

matters is the state of play in one's psychology at that time, and that one is held—and holds oneself—in the right sort of tension in that moment. Stability in resultant dispositions may be sought after and cultivated, but it is not a prerequisite for having the right psychological configuration at a time.

Nonetheless, we do think it an attractive aspect of the dynamic picture that it can make sense of the possibility—should it turn out, on investigation, to be plausible for humans—that stable configurations be developed over the course of a lifetime, in part as a result of an agent's own actions to bring about such a state. Appiah considers only what we might call a “cue-management” conception of how the cultivation of character might be possible:

Given that we are so sensitive to circumstances and so unaware of the fact, isn't it going to be wondrously difficult to develop compassion, say, as a character trait? We just can't keep track of all the cues and variables that may prove critical to our compassionate responses... How, if this is so, can I make myself disposed to do or to feel the right thing? I have no voluntary control on how aromas affect me. I cannot be sure that I will have a free dime show up whenever it would be a fine thing to be helpful. (45-6)

He later offers some positive words about “how to become better people” (71), but almost entirely in terms of engineering one's environment to provide appropriate cues. The diachronic element of the dynamic conception of character, however, offers yet a way in which character development might be possible even with “keeping track of all the cues”. If the situationally-sensitive elements are just one component in a tension system, then the overall product of the other elements could, in principle, be able to wash out or at least drastically moderate the effect of the situational factors. The picture of us as a random conglomeration of transient impulses does not yield that possibility, but the tension system picture does. Again, further empirical investigation would be required to discern whether, and how, such stabilities could be achievable. But our point here is that psychological science does not merely make a case against character, but can also offer ways to see through to a *different* conception of character that can meet our ethical needs while remaining consistent with the situationists' findings.

## The Promises for Intuition

We now turn from Appiah's take on “the case against character” to “the case against intuition”, and again find that we would wish him to travel even further down the naturalist's road than he has allowed himself to journey. Appiah focuses on two possible impacts that experimental science can have on philosophers' uses of intuitions. One we can call a *calibrational contribution*: the science may help us to determine, in cases where our own cognition is conflicted, which of our cognitive inclinations to give credit to. As he writes on p. 94, “[b]eing told that our “intuition” involves the engagement of a different part of our brain from the part that has the reflective idea might just make it easier to side with that reflective judgment (which, let's grant, is sponsored by another set of intuitions).” (See also p. 110, “Understanding where our intuitions come from can surely help us to think about which ones we should trust.”) The other is an *undermining* impact: once science has taught us what the causal processes are that give rise to our intuitions (and also developmental ontogenetic and evolutionary phylogenetic stories of where those processes came from), then we will treat them as *merely* causal entities, capable of explaining and being explained but not of being reasons or justifications. Appiah is cautiously sanguine about the former sort of impact, but is prepared to work a bit of old-school philosophizing to fend off the latter.

Appiah seems worried that the latter, purely negative impact will overwhelm the former, potentially beneficial one. For example, on p. 115 he seems inclined towards an optimism about what experimental results can offer “And so, when we must distinguish the cases where we should and should not take notice of our feelings, we will want to take account of the empirical research at hand. By illuminating odd features of our evaluations, showing where judgments about cases of a certain form diverge from our other, considered judgments, that research can be a useful adjunct to our intuitions, especially if we consider ‘attentive reflection’, as Thomas Reid urged, to be a form of intuition, too.” So far, so good. But he continues, just one page later: “... all this talk of *mistakes* ... has been a distraction from a larger threat.... The lingering unease I mentioned has a different source: the dissonance between viewing moral sentiments as, in some measure,

constitutive of a moral judgment and viewing them as a device that nature has bequeathed us for social cognition.” Appiah tries to manage this dissonance by following Kant in positing a twin set of distinct attitudes, that of *Sinnenwelt* (under which we are “natural beings, governed by causal laws”) and that of *Verstandeswelt* (under which we are agents that act with freedom) (123). And, also like Kant, he finds these attitudes intrinsically incommensurable. For example, he writes that the *Sinnenwelt* perspective is “not a standpoint we can adopt when we ourselves act as agents” (123), and in general the two perspectives are presented as something we have to toggle between. His main thesis in the book is that what we see in the perspective of scientific observation can and must inform what we understand in the perspective of moral evaluation. But there is still, on this picture, a radical separation between them.

Yet this should not be a happy result for Appiah. He spends his introductory pages very persuasively arguing that philosophy and science are separated only by a “waterless moat”, and that continuity, not separation, has been the historical norm. So we are disappointed to find Appiah re-filling that moat with a metaphysical distinction between “stances”. It is a distinction, etymologically at least, between two separate *worlds*, and it is not made clear in Appiah’s book just how, once this distinction is in place, we are to recover that continuity between science’s world and philosophy’s. Now, we too would like to avoid a potentially debilitating undermining impact of science upon philosophy, but we wish to do so in a way that preserves less destructive opportunities for contact between the two as well (and we think that Appiah would wish to do the same).

We would contend that the way to do this is to give the ‘heuristic’ treatment *to that very sense of undermining itself*. That felt conflict between explanation and reason is itself a target for calibration, perhaps sometimes to be trusted, but perhaps more generally to be considered an artifact of the limits of our cognition. For example, Paul Bloom [3] has a set of very interesting studies suggesting that young children may be natural dualists—they seem prepared to treat psychological entities as not needing to obey the same basic rules of physics that they expect to apply to blocks and toys. For example, the children seem to expect that physical objects cannot jump across space, but must move in continuous paths through it; whereas psychological entities, like persons, can

appear in one spot and then in another without ever occupying any intervening positions. Bloom offers the explanation that the children have two different cognitive systems trying to tell them what to expect, and they cannot yet apply more than one system at a time to any given object. So, if you’re a person, and thus the ‘theory of mind module’ should apply to you, you can’t also at that time be an object, such that the ‘theory of body interaction’ module can apply to you. One need not view this as a deep philosophical commitment, but one can see it as just a resource limitation—a performance error—that we all come to be able to overcome even before puberty. But, even if we all can learn how to let *some* of our distinct cognitive systems literally co-operate, perhaps for other systems this is not so easy. Perhaps it is just a greater resource limitation that makes it hard for us to keep both the parts of our cognition that handle explanations, and the parts that handle justification, working together at the same time. This may be experienced as an undermining, for whenever the explanation-giving machinery becomes strongly active, we may typically feel the justification-evaluation machinery dampening. But we don’t *need* to treat it as such. Perhaps we can find a way of viewing the world as possessing both explanations and justifications all at once, in one-but-not-entirely-unified perspective.

We would further suggest that naturalism, properly understood, enables us to see the potential contributions of science as ones that can happen *internal to, and not merely prior to*, the process of philosophical deliberation. We take it that a kind of abductive (or perhaps quasi-abductive) mode of argument has become fairly standard in philosophy, in which a range of claims are put forward as *prima facie* and *pro tanto* considerations against which the success of different theories are to be measured. If part of what philosophy must do is *tithemi ta phainomena*, then one role that science can play is thus as providing some of those *phainomena*. But science can also help evaluate the costs of giving up on claims that a particular philosophical view might try to give up. For example, if a given claim is found universally, it may reasonably be considered a more important piece of phenomena to preserve than another claim that displays a fair amount of variation.

The sciences can make a still deeper contribution to ongoing philosophical deliberation by offering us better illumination as to where the joints of nature –



including human nature – are for the cutting. A number of standard philosophical modes of argumentation require an operational sense of the contours of the world. A slope's slipperiness is a matter of whether it offers any hand-holds along the way. For example, some participants in debates over abortion have in the past made use of the idea of quickening as a possible place to stop a slide of personhood somewhere between post-natal infant and fertilized ovum. Advances in developmental biology have brought with them a range of other candidate slide-stoppers, such as the capacity to experience pain (typically not until the third trimester). Such advances do not settle the debate, and, again, let us emphasize that we are not advocating a form of naturalism that attempts a wholesale replacement of philosophy with science. But such advances do enrich philosophical inquiry, and make positions intellectually available that had not been so before.

Or consider a highly pertinent epistemological topic: the methodological trustworthiness of intuitions in philosophy. George Bealer, in his strenuous defenses of intuitions, has offered several arguments that turn crucially on a particular construal of epistemic natural kinds [4]. This licenses, for him, an induction from our apparent intuitive success in a vast range of ordinary categorizations to the reliability of intuition in the more far-flung cases of the sort more typically deployed in philosophical argument. Yet some recent psychological has raised doubts about the idea that our intuitive judgments form one seamless whole. Our minds may be more of a "Swiss army knife" (E.g., [5]), an assemblage of a perhaps very large number of specialized processors; Appiah has a nice discussion, following Haidt, of significant modularity just within our moral evaluations, let alone within our cognition at large. If that is so, then it may not be possible to run an induction from our epistemically successful performance on one set of cases to a hoped-for success on others. It would be like inferring the sturdiness of one's hammer from the sharpness of one's saw—they may both come in the same toolbox, but beyond that they may not form an appropriately projectible kind.

Bealer has also argued, among others (e.g., Sosa)<sup>3</sup> that a radical empiricism that would deny positive

epistemic status to intuitions that it would nonetheless grant perception, is guilty of a vicious form of ad-hocery. Viewed in the abstract, this argument can seem compelling—why hold these different practices to different standards, unless out of sheer scientific chauvinism? But one of us has argued [6] that, as a matter of fact, when one attends to the history of science and to the nature of our actual perceptual and scientific practices on the one hand, and our philosophical practices with appeals to intuition on the other, salient differences emerge. In particular, the former possess a great deal of what is termed "hopefulness": resources for the detection and correction of errors as they arise in investigation. Yet the former can seem somewhat ... hopeless, at least at present. When intuitions clash, we possess little means for adjudicating the dissensus. And, it is argued, this difference in hope makes an important epistemological difference, and gives us reason to trust our practices that are hopeful and not trust those that are not. It may well be, then, that attention to science (and, in this case, the history of science) may make a distinction that seemed ad hoc at one point in an investigation, become a pertinent point of cleavage at a later one. Ironically, the epistemic "hopelessness" of armchair intuitions, according to this argument, is somewhat due to its estrangement from science, and in particular the psychological science of the sources of the intuitions. Such science would exactly help us in the calibrational ways discussed by Appiah, and thus provide us exactly with some of the necessary resources to help preference some one out of a set of conflicting intuitions. So the practice of appealing to intuitions may yet become more hopeful—but only if it first becomes more friendly to naturalism.

## Conclusion

It is a virtue of Appiah's *Experiments in Ethics* that its author is honest in reporting his conflicted intuitions about the relation between a naturalistic epistemology and ethical insight. We have tried to make the case here that this conflict is nonetheless not a necessary one, and that a more fully embraced naturalism can also offer the promise of a richer, more fully satisfying ethics. Whether this promise can be fulfilled will be a matter of both further empirical investigation and further armchair reflection, and to the extent that these branches

<sup>3</sup> Bealer, G. "The Incoherence of Empiricism"; Sosa, E., "A Defense of Intuitions"

can twine each around the other, the stronger a stance we will have for when we have to go out on a philosophical limb. One sees the world in its fullness better with the perspectives of both eyes open at the same time.

## References

1. Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics*, 114: 458–491.
2. Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
3. Bloom, P. (2004). *Descartes's Baby: How the Science of Cognitive Development Explains What Makes Us Human*. New York: Basic Books.
4. Bealer, G. (2000). A Theory of the A Priori. *Philosophical Quarterly* 81: 1–30.
5. Pinker, S. (1997). *How the mind works*. New York: W. W. Norton.
6. Weinberg, J. (2007). “How To Challenge Intuitions Empirically Without Risking Skepticism,” *Midwest Studies in Philosophy*. 318–343.