

Why Character Traits Are Not Dispositions

Roderick T. Long

Auburn University | Molinari Institute | aaeblog.com

Abstract: While it is customary to think of character traits as dispositions, I argue that this is a mistake. Character traits have dispositional features, but are not exhausted by those features. Were they so exhausted, it would not be possible for dispositionally identical people to differ in their character traits; but I argue that if character traits describe what people are like morally, then such a divergence will be possible on both deterministic and indeterministic accounts of agency. In closing I discuss the implications for Harman's "attribution error" thesis.

It is customary to think of character traits as dispositions – specifically, dispositions to perform certain sorts of actions in certain sorts of circumstances. I shall argue that this is a mistake. I do not mean to deny that character traits have dispositional features; rather, I deny that character traits are *exhausted* by their dispositional features. Were they so exhausted, it would not be possible for dispositionally identical people to differ in their character traits; but in fact we have good reason to grant that possibility.

I shall proceed by first examining why the notion of character traits as dispositions initially seems attractive, and then considering in turn three hypotheses about human action: that it is not causally determined; that it is causally but not psychologically determined; and that it is both causally and psychologically determined. I shall argue that under each hypothesis we have reason to grant the possibility of dispositionally identical people differing in their character traits. I shall close with some remarks on the potential relevance of my conclusions to the debate over Gilbert Harman's "fundamental attribution error" thesis.

The chain of reasoning that leads to regarding character traits as dispositions is natural enough. If we want to know what someone is like morally, we first consider what she does. But then we notice the role of opportunity, or its lack, in explaining a person's actions: perhaps she did nothing especially courageous because the relevant challenges never arose; or perhaps she did nothing especially dishonest because she had no chance of escaping detection. And so we come to recognise that what a person *would* do in various circumstances is crucial to defining her character – whereupon what she *actually* does seems to recede to the status of a mere subset of what she would do. Yes, she rescued that crocodile from the ravenous baby, but that's because rescuing crocodiles from ravenous babies is just what she *would* do in circumstances like those obtaining; and someone else who likewise *would* do it, but was never placed in the right circumstances, is still the same kind of

person as she is. Determine what a person *would* do in various circumstances, and you've settled what she is like morally; once we know that she would do X in circumstance Y, seeing her *actually* doing X when placed in circumstance Y gives us no new information. And so our evaluation of character shifts from the indicative to the subjunctive, from the actual to the counterfactual.

That our view should *expand* from the actual to *include* the counterfactual I readily grant. But that it should *shift* to the counterfactual, no longer giving any independent weight to the actual, I deny. To see why, let's consider how actual and counterfactual are related under the various relevant hypotheses about human action. There are initially four possibilities:

- a) human action is **deterministic**, and its being so is **compatible** with moral agency;
- b) human action is **deterministic**, and its being so is **incompatible** with moral agency;
- c) human action is **indeterministic**, and its being so is **compatible** with moral agency;
- d) human action is **indeterministic**, and its being so is **incompatible** with moral agency.

Defenders of (b) maintain, while defenders of (a) deny, the *incompatibilist* thesis that an action's being causally necessitated by prior factors running back beyond the agent's volition renders the agent unable to do otherwise in the manner required for moral agency – which, combined with determinism, yields *hard determinism*. Defenders of (d) maintain, while defenders of (c) deny, (what I call) the *inverse-incompatibilist* thesis that an action's being causally undetermined deprives the agent of such control over her actions as is required for moral agency – which, combined with indeterminism, yields *hard indeterminism*.¹

But the first two options can be further precisified, given the distinction between general causal determinism and *psychological* determinism – the latter being the thesis that every action has temporally antecedent sufficient conditions in the agent's mental states. Since mental states are multiply realisable, two psychologically indistinguishable agents might nevertheless respond differently to the same stimuli, in virtue of differences not in the prior mental states themselves but rather in the underlying physical realisers of those states. Thus psychological determinism might be false of human action, even if causal determinism is true of it (though not vice versa).

¹ My own view is that incompatibilism and soft indeterminism are true, while compatibilism is false. But my argument in this paper does not depend on which view is true.

Hence we need to distinguish six possibilities in total:

- a1) human action is **psychologically deterministic**, and its being so is **compatible** with moral agency;
- b1) human action is **psychologically deterministic**, and its being so is **incompatible** with moral agency;
- a2) human action is **causally** but **not psychologically deterministic**, and its being so is **compatible** with moral agency;
- b2) human action is **causally** but **not psychologically deterministic**, and its being so is **incompatible** with moral agency;
- c) human action is **indeterministic**, and its being so is **compatible** with moral agency;
- d) human action is **indeterministic**, and its being so is **incompatible** with moral agency.

But the traditional concept of character traits has its home in the context of moral agency – is a moral or ethical concept – so any (non-debunking) analysis of character traits must take the reality of moral agency as a premise. So if we are determinists and want to take moral character seriously, we must be compatibilists and soft determinists (though we need not be inverse-compatibilists); and if we are indeterminists and want to take moral character seriously, we must be inverse-compatibilists and soft indeterminists (though we need not be compatibilists). Hence for present purposes we can dismiss (b1), (b2), and (d), leaving (a1), (a2), and (c) – which I’ll take up in reverse order, since the point I want to make is more easily seen in (c) than in (a2), and more easily seen in (a2) than in (a1).

Hypothesis I: Indeterministic Moral Agency

Suppose human action is indeterministic. In that case, dispositions of character will be merely probabilistic, and will be consistent with more than one outcome even given the same situation. Suppose further, then, that Abel and Mabel are dispositionally identical, and neither has ever committed a murder or had any real opportunity to do so, but each is placed in a situation in which the probability of committing a murder is 65% – whereupon Abel commits the murder but Mabel does not. Suppose still further that Abel and Mabel remain dispositionally identical after the murder. (The latter supposition may seem unrealistic, since one’s actions generally have an impact on one’s dispositions; but it does not seem impossible.)

It would be a mistake, I think, to infer that Abel and Mabel are also identical in respect of character. Abel had a chance to commit murder, and he took it. Mabel had just the same opportunity and just the same motive, but she refrained. So if character is a matter of *what someone is*

like morally, it's hard to say that Abel and Mabel are alike in character. Nor can we argue for Mabel's moral identity with Abel by saying that she *would* have acted like Abel if placed in Abel's circumstances, because she *was* in just those circumstances and did *not* act like Abel. Their moral profiles are different even though their dispositions are alike.

It may be objected that on the indeterminist hypothesis, it is a matter of sheer chance whether either Abel or Mabel commits the murder or not, making the eventual difference between them objectionably morally lucky and so not a proper basis for a difference in what the two are like morally. But this objection would be out of place. Recall that we have dismissed the three hypotheses (b1), (b2), and (d), on the grounds that in denying moral agency, they all deny a presupposition of any non-debunking analysis of moral character. Our indeterminist hypothesis is (c), not (d). If being undetermined rendered Abel's and Mabel's decisions objectionably morally lucky, it would deprive them of moral responsibility, which would fit (d), not (c). By accepting (c) as our hypothesis, we have committed ourselves to granting either that actions' being probabilistic is not a matter of moral luck, or else that if it is moral luck it is not of the sort that imperils moral agency.² Thus the objection is estopped from the outset.

Hence if human actions are indeterministic but moral character is real, then character traits are not merely dispositional.

Hypothesis II: Causally But Not Psychologically Deterministic Moral Agency

Let's turn to the next hypothesis, that human actions are causally but not psychologically determined. In that case, Abel and Mabel may still react differently in the same circumstances even if they are *psychologically* identical, so long as there is some causally relevant difference in the underlying physical states and processes in which their identical psychologies are realised. In such a case, are Abel and Mabel dispositionally identical? The question is ambiguous. If we confine ourselves to *psychological* dispositions, they are; if we have in mind the full range of dispositions, physical as well as psychological, they are not. But in neither case can we identify character traits with dispositions.

Character traits cannot be identified with narrowly psychological dispositions, because Abel and Mabel are alike in respect of their psychological dispositions, yet we have exactly the same reason for

² I have argued elsewhere that the first option is correct, *i.e.*, that indeterministic agency is not properly described as chancy or random; see "Free Minds and Future Contingents," online at: <http://praxeology.net/TulaneFreewill5.pdf>

regarding them as different in character as we did under Hypothesis I – namely that each had the same opportunity and motive to commit murder, but one refrained while the other did not. Nor can character traits be identified with dispositions in the broad sense, because if psychological states are multiply realisable, character traits must be so as well, on the same grounds.

It may be objected that if the difference between Abel and Mabel turns not on their psychologies but solely on differences in the underlying microphysical states and processes constituting those psychologies, then that difference is objectionably morally lucky and so not a proper basis for a divergence in moral character. But this objection is ruled out for the same reason that the analogous objection was ruled out under Hypothesis I. Our Hypothesis II is (a2), not (b2); but if being the results not of differences in psychological states but of differences in the physical realisers of those states rendered Abel's and Mabel's decisions objectionably morally lucky, it would deprive them of moral responsibility, which would fit (b2), not (a2). By accepting (a2) as our hypothesis, then we have committed ourselves to granting either that the difference between Abel's and Mabel's actions is not a matter of moral luck, or else that if it is moral luck it is not of the sort that imperils moral agency. Hence if human actions are causally but not psychologically deterministic, and moral character is real, then character traits are not merely dispositional.

Hypothesis III: Psychologically Deterministic Moral Agency

Let's turn, finally, to the possibility that human action is not merely causally but even psychologically deterministic. The case for the possibility of divergence in character traits between dispositional identical agents is harder to sustain under this hypothesis than under the first two – though as this is the least plausible of the three hypotheses I don't find that too worrying. Still, let me try to make the case.

Suppose once again that Abel commits a murder and Mabel does not. Under the present hypothesis, it follows that either Abel and Mabel differed in their psychological dispositions, or else they faced different circumstances.

Consider the latter option first: Abel and Mabel are identical in their psychological dispositions, but Abel committed murder and Mabel did not because they were placed in different circumstances. In this case we must say, as we could not say under Hypotheses I and II, that if Mabel had been in Abel's circumstances, she would have committed murder as well. But does that make them morally alike? Steam and ice have, at a deep level, the same dispositions (including the disposition to manifest as steam in some circumstances and as ice in others), yet they are not physically alike.

Steam is hot and ice isn't. Of course ice *would* be hot if it were heated up enough, but steam is hot *now*. Why should we not likewise say that Abel differs morally from Mabel because Abel *is* a murderer? True, Mabel *would* be a murderer if placed in the same circumstances, but Mabel is a murderer *now*; the fact that they *would* be alike morally under conditions different from those actually obtaining does not make them alike morally now. (If the divergence between the two seems too objectionably morally lucky to ground a difference in moral character – well, we already ruled out (b1) in favour of (a1); by now you know how the rest goes.)

Consider, next, the other option: that Abel and Mabel faced the same circumstances but acted differently because they differed in their psychological dispositions. Nevertheless, it is possible that they are identical in their psychological dispositions *now*, post-murder. Sometimes the effect is to reinforce the disposition that produced the action; but sometimes not. Suppose that Abel was initially more inclined than Mabel to commit murder (which is why he did), but the shock of actually committing murder has put him off it, so that he is now more averse to murder than he was before – indeed, he is now exactly as averse to murder as Mabel already was. So the two are alike in psychological dispositions, but Abel is a murderer and Mabel is not, so they are not morally alike – hence they differ in moral character despite having the same dispositions.

Objection: the plausibility of saying that Abel's being a murderer makes him morally unlike Mabel turns on the assumption that a murderer is *worse* than, and so morally unlike, a non-murderer. But in order to imagine Abel and Mabel as dispositionally identical, we had to make Abel a *repentant* murderer, and it's not clear that a repentant murderer is worse than a non-murderer with exactly the same dispositions.

Reply: my argument doesn't turn on a murderer's being *worse* than a murderer, only on their being morally *unlike*. The character of anything is what it's like; and the moral character of a person is what she's like morally. What's Abel like morally? He's (*inter alia*) a repentant murderer. We can't say the same about Mabel. So even on the assumption of psychological determinism, dispositional identicals can differ in character traits.

Thus whether we assume Hypothesis I, Hypothesis II, or Hypothesis III, character traits are not dispositions. And those three hypotheses are the only ones consistent with granting the existence of moral character at all. Hence if we are to take moral character seriously, we cannot regard the traits that compose it as being purely dispositional. (Admittedly, as noted above, while I think my argument under Hypothesis III is correct, I don't take it to be as compelling as my arguments under

Hypotheses I and II; but, again as noted above, Hypothesis III seems so unlikely that I am happy to rest my case on I and II alone if need be.)

The “Attribution Error” Error

So what are the implications of these findings for the “attribution error” debate? Gilbert Harman maintains that experiments in social psychology give us reason to doubt or deny the existence of stable dispositions for action, *and so* of character traits.³ His argument evidently assumes, then, that character traits are dispositions; without this identification, the argument for the nonexistence of character traits does not go through.

I don’t think we *need* to make this point to refute the “attribution error” argument; it has plenty of other problems even if we assume that character traits are dispositions. This is not the place for a full discussion of the issue, but let me note four problems in passing:

1. The fact, should it be one, that the best predictor of people’s actions is their circumstances does not show that they lack stable dispositions for action, any more than the fact that the best predictor of whether a match will catch fire is whether it is dry or wet, and struck or unstruck, shows that matches lack stable dispositions for catching fire. A disposition and its circumstances of manifestation are not *competing* explanations of its manifestation. At best, the experiments Harman cites show only that people have the *same* stable dispositions for action, not that they lack such dispositions.⁴
2. But in reality they don’t show even that much, since most of the experiments reveal a greater-than-zero degree of variation among the subjects – thus *confirming* rather than *disconfirming* the existence of differences in stable dispositions. What the experiments demonstrate is a *higher than expected* amount of uniformity, not *total* uniformity. (Also: higher than expected, *by whom?* But let’s leave that question for point 4.)
3. A character trait, as traditionally understood, is a disposition, yes, but a disposition not merely for overt behaviour but for *action* – where action is behaviour that is done for a *reason*. Hence two actions that are identical in terms

³ Gilbert Harman, “Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error,” *Proceedings of the Aristotelian Society* 99 (1999), pp. 315-331; cf. Gilbert Harman, “The Nonexistence of Character Traits,” *Proceedings of the Aristotelian Society* 100.1 (June 2000), pp. 223–226.

⁴ For elaboration of the argument that Harman’s case, if successful, would show not that people lack stable dispositions but rather that they have the same stable dispositions, see Jennifer McKittrick, “Defending Your Character: A Reply to Harman” (unpublished, but presented at the Alabama Philosophical Society, Orange Beach, 24 October 2003).

of overt behaviour will still belong to different action types if they are done for different reasons. Hence in order to show that people do not differ in their dispositions for action, one would need to test not just for differences in overt behaviour but also for differences in *reasons*; but most of the experiments that Harman cites either do not do this, or do it in only the most cursory of fashions.

4. If one wishes to subject the traditional understanding of character traits to empirical test, one must first correctly identify *what* predictions that traditional understanding yields. It's of little use to spend one's time triumphantly falsifying predictions if those predictions are not the ones that the view to be refuted actually generates. Most of the literature that Harman cites falls afoul of this problem. Consider just two instances.

The Good Samaritan experiments evidently assume that a traditional understanding of character traits must predict that virtuously charitable subjects will stop to help others even when they are hurrying to a prior obligation, while those lacking the virtue of charity will not stop to help even when they have plenty of time. But the duties of charity are traditionally *imperfect* duties, that is, duties that do not call for the corresponding action to be performed at every possible opportunity, but only *regularly*, thus leaving the choice of occasions for performance largely up to the agent's discretion; hence a virtuously charitable agent's letting the presence or absence of a prior obligation determine whether to provide assistance on a particular occasion is not unexpected. Likewise a person who lacks the virtue of charity may nevertheless provide help when the cost of doing so is low – perhaps to win others' approbation, or ease her own conscience, or even out of genuine other-regard that falls short of full charity, so the fact that those with and those without the virtue of charity would behave similarly in the experimental setup is not unexpected. A true test would turn on the performance of a *perfect* duty.

In this regard the Milgram experiment is more apt, since it does concern a perfect duty. But in order for the Milgram experiment to pose a problem for virtue ethics, it must be assumed that its results are different from what would be expected on the traditional view. Yet virtue theorists have been complaining ever since Socrates that most people are excessively conformist, unstable in their convictions, and too concerned with looking good to others, while more recent virtue theorists have often bemoaned the widespread tendency of deference to authority; so the Milgram results are not all that unexpected. (Nor is a widespread tendency the same as a universal tendency; if Milgram had drawn his subjects from a group of anarchists, I suspect the results would have been rather different.)

Thus even if character traits were dispositions, no implication as to their uniformity, let alone their nonexistence, would follow. In fact, however, as I have argued, character traits are not purely dispositional, since dispositional identicals can be different in character. Hence even if it were true, as per Harman's arguments, that stable dispositions for actions are either uniform or nonexistent, it

would not follow that *character traits* are uniform or nonexistent. What people are like morally depends on what they actually do, not just on what they would do.