

What Makes a Theory a Theory of Virtue?
Comments on Spelman's "Consequentialist Theories of Virtue"

Sara Worley
Bowling Green State University

Spelman suggests we can resolve the debate between Driver and Bradley by noting that there are at least two different senses of "virtue", $\text{virtue}_{\text{fact}}$ and $\text{virtue}_{\text{expectable}}$, which can be modelled on Parfit's distinction between various senses of "right" and "wrong." $\text{Virtue}_{\text{fact}}$ comports with Driver's actualist understanding of consequentialist virtue, where a trait is virtuous just in case it leads to good consequences. $\text{Virtue}_{\text{expectable}}$ suits Bradley's discussion, where what matters for determining some traits virtue are its expected consequences. So we needn't decide between the two views; we can have both.

I do think there's something to the claim that we ought accept both $\text{virtue}_{\text{fact}}$ and $\text{virtue}_{\text{expectable}}$. The distinction seems somewhat analogous to the distinction between truth and justification. Truth is really the target of our belief formation practices, and ideally we would like behavior to be informed by and responsive to true beliefs. Given our epistemic limitations, we are often not in a position to know what's true, and have to settle for having our practices and behavior be informed by (merely) justified beliefs. We of course hope that our beliefs are true, but we can't know that they are. Similarly, consequentialists might want to have, and for others to have, character traits which in fact lead to good outcomes, but, given our epistemic limitations, we can't know in advance which these are. So the best we can do is inculcate traits which we think will lead to good outcomes. So just as the distinction between truth and justification is important, and we need both, we may need both $\text{virtue}_{\text{fact}}$ and $\text{virtue}_{\text{expectable}}$.

Nonetheless, I do have a couple of concerns. The first is whether, even given the argument which I've just sketched out, the $\text{virtue}_{\text{fact}}$ sense doesn't stretch the meaning of "virtue" a little too far. First note that Spelman seems to be relying on what Bradley calls an individualist sense of virtue. According to it, some trait T is a virtue for S if the world would be better off if S exhibited T. This contrasts with a universalist sense, where we evaluate whether T is a virtue by asking whether the world would be better off if everyone exhibited it. For example, we might think that honesty is a virtue because the world as a whole would be better off if everyone were honest.

Nonetheless, we might think that honesty is not a virtue for a particular diplomat, who needs to engage in some strategic deception in order to achieve his (otherwise admirable) diplomatic goals. Spelman is relying on an individualistic sense, since he focuses on, e.g., the actual and expected results of Helper's generosity, not on the actual and expected results of generosity in general.

Instead of thinking about Helper, let's think about Lucky. As Bradley characterizes him, Lucky is a murderous psychopath. He never manages to kill anyone, because of massive amounts of moral luck, e.g., sudden gusts of wind displace bullets, birds get in the way, etc. He never actually succeeds in doing any harm, but he doesn't do any good, either. But let's change the case a little. He still doesn't do any harm (suppose that most of his targets don't ever find out that they were shot at, so there aren't even any psychological costs). However, one of his prospective targets does become aware, and this causes him to take some further action (leaving town, barricading his home) which turns out to, in the actual world, have a number of good consequences, which would not have happened had Lucky not tried to kill him. Would we then want to say that murderous psychopathy was a virtue for Lucky? It seems that we would, on Spelman's account. His being a murderous psychopath had a net positive balance of good over evil, so it's a *virtue_{fact}* even if not a *virtue_{expectable}*.

This seems an unattractive result. Whether a trait is a virtue or a vice should not depend, it seems, merely on moral luck.

One might think that the difficulty here has to do with the distinction between individualist and universalist accounts. It certainly seems true that a world in which everyone is a murderous psychopath would be worse than a world where no one is, so we can say that it's a universalistic vice, even if an individualistic virtue. But I think there might be something else going on here. Note first that although I've been following Spelman in interpreting "expected consequence" as an epistemological category, it is not for Bradley. "Expected consequence" for Bradley is modeled on the notion of "expected utility" where one computes an expected outcome for an event by summing the products of the probability of various possible outcomes of an action by the goodness of those outcomes. The expected consequence of Lucky's behavior are quite low, even though the actual consequences are good, since he does not benefit from massive moral luck in most of the relevantly similar worlds in which he attempts to kill people. Crudely, the trait has a bad expected outcome because it leads to bad outcomes in the vast majority of relevantly similar possible worlds.

One might think this is still just a matter of epistemology, since expected utility is relevant under conditions of uncertainty. If we knew what the outcomes of our actions would be, we would not need to calculate expected utilities. Similarly, given that we know what the actual outcomes are, we needn't worry about expected outcomes. But I think there's more to be said than that. Virtues and vices are character traits. They are dispositions. To attribute a disposition to someone or something is to do more than just sum up a set of observed behaviors. It's also to commit oneself to counterfactual claims about which behaviors would be exhibited under various circumstances. If we attribute solubility to salt, for instance, we are making claims about it would do if placed in water. The truth of those claims are independent of the actual history of a piece of salt. It doesn't cease to be soluble just because it is never in water. Similarly, to say that a person has a character trait is not simply to sum up her past behavior. It's also to make claims about how she would have behaved in various other, non-actual, circumstances. It seems, then, that to say that one ought to have had a certain disposition, is to say more than that the behaviors one actually produced happened to have good consequences. It's also to say that the behaviors one would have produced in other circumstances would also (likely) have had good consequences, all other things being equal. Furthermore, there ought to be a particular explanatory relationship between the good outcomes and the trait. The trait is a good one to have precisely because it's the kind of thing which (reliably) generates good outcomes in the sorts of circumstances which we can expect to encounter, not one which merely, and luckily, happened to generate good outcomes in the situations we did encounter.

Something like this is, I think, at work in the account of the lying diplomat who is nonetheless virtuous. This diplomat's deception can be regarded as virtuous precisely because it's skillful, targeted, and strategic. He knows how to lie appropriately, and with skill, and in a way which is liable to (given facts about human psychology and the circumstances which he encounters) produce the results that he wants. And, part of what it means to say that he's skillful, is to say that, were the circumstances different, he would be able to change his behavior appropriately, so as to meet the new circumstances. Another component is that there's a reliable, rather than lucky, causal connection between the lies and the outcomes. An inept diplomat who lies clumsily would not be regarded as virtuous in virtue of his clumsy lies, even if they happened, by sheer luck, to produce good results.

The suggestion, then, is that, for a virtue consequentialist account, we need to be focused on whether someone ought to have, or have had, a particular character trait, not just whether a particular set of behaviors

happened (accidentally or otherwise) to have good outcomes. If all we can show is that Lucky's attempt to murder one particular person had a good outcome and the rest of his behavior was neutral, then at most what we would have shown was that one piece of behavior was good. We would not have shown that Lucky ought to have had the character trait of being a murderous psychopath. To show that, we'd need to show that being a murderous psychopath is the sort of trait which is liable to produce good outcomes, in the variety of circumstances that Lucky was likely to encounter. That is, we'd have to show that its expected outcome (in Bradley's sense) is high, regardless of the actual effects of those traits.

If this is right, would it mean that we have to give up on the actualist/expectable distinction? I don't think so. A virtue consequentialist can still make the distinction between traits that actually tend to produce the right outcomes in a variety of circumstances, and traits which we merely think tend to produce the right outcomes in a variety of circumstances. That distinction still seems a good one. The suggestion is, instead, that if we are to take the notion of virtue and vice seriously, we ought focus not just on the actual effects of actually occurring behaviors, but on the slightly broader question of which behavioral traits would tend to produce the right outcomes in a variety of situations.