Response to my Critics

JULIA DRIVER
Dartmouth College

This essay is a rejoinder to comments on *Uneasy Virtue* made by Onora O'Neill, John Skorupski, and Michael Slote in this issue. In *Uneasy Virtue* I presented criticisms of traditional virtue theory. I also presented an alternative – a consequentialist account of virtue, one which is a form of 'pure evaluational externalism'. This type of theory holds that the moral quality of character traits is determined by factors external to agency (e.g. consequences). All three commentators took exception to this account. Therefore, the bulk of my response focuses on defending the externalist account of virtue presented in the final chapters of *Uneasy Virtue*.

I am very appreciative of the time and effort my commentators took in their discussions of *Uneasy Virtue*. Each commentator has pushed me hard to justify the position that I argue for in the book. While I disagree with many of the comments, others have forced me to refine and clarify my view.

In *Uneasy Virtue* I had two very broad aims. The first was to argue that the classical view of virtue, particularly the view presented by Aristotle, which has had a very large influence over modern developments in virtue ethics, is mistaken. I argue this by showing that it has counter-intuitive results – it places too many psychological requirements on the possession of virtue. This results in there being very few who would qualify as virtuous, and would rule out many cases of virtue which we tend to think of as plausible cases of virtue. One example I use to motivate this claim is that of modesty. I argue that a modest person underestimates herself in some respect or other, to a limited degree. This means that modesty involves epistemic defect, and thus would not be allowed as a virtue on the classical account. I then argue that what makes a trait such as this valuable are the effects it produces – and that this suggests a general strategy for developing an account of virtue along consequentialist lines. And this is the second broad aim of the book – to present such an account. I argue that the best account is one which is an example of pure evaluational externalism. That is, the moral quality of the character trait is determined solely by factors external to agency, i.e. consequences. This view is to be contrasted with pure evaluational internalism as well as what I call a 'mixed' view. The former holds that the moral quality of the character trait is completely determined by factors internal to agency, i.e. intentions, or motives. The latter holds that while such internal states are necessary, they are not sufficient – what is also needed is a connection to what actually occurs in the world. On my view a writer

© 2004 Cambridge University Press
DOI: 10.1017/S0953820803001043
Utilitas Vol. 16, No. 1, March 2004
Printed in the United Kingdom

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

like Kant is arguably an internalist with respect to moral evaluation, whereas Aristotle offers us a mixed view.

Professors O'Neill and Skorupski both focus on the internalist/externalist distinction that the account relies on. It is the mixed view, I suspect, that most people find intuitively plausible. However, one claim in *Uneasy Virtue* is that this mixed account is the one which, in theoretical terms, is also the most problematic. This is because the mixed view inherits problems from each of the 'pure' alternatives. O'Neill believes that I classify the Aristotelians as internalist, but that is not entirely true: the Aristotelians, I believe, best exemplify the mixed view. That is, a view which regards internal states as necessary but not sufficient, since virtue must also, in some way, lead to human flourishing.

O'Neill rightly points out that on my view we could well be mistaken in our conventional judgements of virtue. The example I use is that of generosity: suppose that it turned out that generosity to others actually undermined their character in some deep way that could not be counteracted effectively. Then, on my view, it turns out not to be a virtue. O'Neill cites the case of Ayn Rand, and her view that selfishness is indeed a virtue. But of course it is not enough to show that sometimes selfishness has good effects. I believe Rand to be quite mistaken. The good effects must be systematic and outweigh the bad effects also produced by the trait. So, while it is indeed possible that selfishness could turn out to be a virtue, the empirical claims necessary to support this seem wildly implausible to me.

Thus, while I agree that on my account there could in principle be a radical transformation of our list of virtues and vices, this seems quite unlikely. The actual account is not all that radical – it is meant to be no more radical than Hume's, and follows his suggestion that our judgements of virtue are transformed (as they should be) when bad (or good) effects of a trait are realized.[1]

O'Neill also faults me for a possible inconsistency. She notes that I seem to propose 'that we limit the traits that are to count as virtues to those that produce *intentional* action ... I am not sure where this leaves modesty and other virtues of ignorance'. But, in fact, my claims about intentional action and the virtues of ignorance are not

[1] For example, in *An Enquiry concerning the Principles of Morals*, section II, part II, Hume writes: 'whenever disputes arise, either in philosophy or common life, concerning the bounds of duty, the question cannot, by any means, be decided with greater certainty, than by ascertaining, on any side, the true interests of mankind. If any false opinion, embraced from appearances, has been found to prevail; as soon as farther experience and sounder reasoning have given us juster notions of human affairs; we retract our first sentiment, and adjust anew the boundaries of moral good and evil' (*An Enquiry concerning the Principles of Morals*, ed. J. B. Schneewind (Indianapolis, 1983), p. 19).

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

at all incompatible. The reason I tie virtue to intention is that I believe one needs to be the sort of being who performs intentional actions, or at least formulates intentions, to be virtuous. However, the agent in being virtuous need not be intentionally acting under the 'virtuous' description of the action. So this is entirely consistent with my account of the virtues of ignorance. Jane Bennett intentionally defends Mr Bingley's sisters to Jane; what she does not intentionally do is overestimate their good qualities. What I wanted to do is offer a way of distinguishing the good-producing traits of things like robots, from those of persons. I stand by this account, though I do think that it could be supplemented by some material on autonomy, for example. As Nomy Arpaly points out in her recent book, one sense (of very many senses) of autonomy that is commonly bandied about ties autonomy to responsiveness to reasons.[2] And responsiveness to reasons need not involve a conscious and explicit deliberation using those reasons. But it will involve the agent being the sort of being that acts intentionally: no other beings are responsive to reasons. To have moral virtue one must be a moral *agent*, the sort of being exemplifying *agency*, and intentional action.

But O'Neill has deeper misgivings about the account that I offer. She is quite right to ask for clarification of my account on the issue of internal states vs. dispositions. O'Neill disagrees with my view that the internalist view is the least problematic, and also disagrees with my, albeit tentative, identification of Kant as an internalist in this regard.

I certainly defer to O'Neill on the correct interpretation of Kant. However, I would like to point out that the claims I made were quite tentative ones and really intended to spell out a possible interpretation, and one which I think does, to many, have a good deal of plausibility. Casting Kant as an evaluational internalist helps to explain his view that on his account moral luck issues are not as serious as they would be for those who feel that the actual deed (and not the internal state), or what is accomplished, needs to be good in order for the act to be good. Indeed, I present this as a considerable advantage of this approach. One problem for the externalist account will be to try to find some way to work around this problem, or some plausible rationale for simply biting the bullet on it.

O'Neill believes the positive account that I offer is flawed because it assumes a robust view of character traits as knowable, and knowable independently of what they bring about. She makes a distinction between a robust account or a minimalist/ascriptive account of character. The robust account is causal; the ascriptive account is

[2] See Nomy Arpaly, *Unprincipled Virtue* (New York, 2003), ch. 4.

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

not. The robust account commits one to the metaphysical view that character traits in fact exist; the ascriptive view does not commit one to the existence of a specific character trait (since this will be underdetermined by the evidence). I think that O'Neill is right to take me to task over the causation issue. If I were to rewrite the account I would not commit myself to strict causation between the traits and the 'effects' – but I also think that consequentialism need not be spelled out in such a way that we say x causes effect y, and it is in virtue of effect y that x is morally significant. The reason one is tempted to think strictly in terms of causation is that these traits 'make a difference' somehow, and it is easy to slip from 'making a difference' to causation – but of course, this need not be the case. So, what is crucial to my account is that there are these character traits, and they do make a difference, and the ones that make a positive difference are the virtues. This, I think, qualifies as a minimal account. I cannot see how non-consequentialists themselves could do without it, at least non-consequentialists who also believe there are such things as virtues (I take Kant to be one of these). If one is not committed to this, then why do we try to engage in moral education? What is it that we are doing when we try to inculcate character traits which are good ones? I do not believe virtues are simply useful fictions. But if I did it would be on the basis of some other consideration, such as the empirical evidence provided by some social psychology experiments.

We must be able to infer character traits, or to think that character traits (virtues) are the sorts of things that make a difference (whether we want to understand that in strictly causal terms or not); if character traits are not knowable, if they do not make any difference, then how is it that we hold people at all responsible for them? If her point is that they are knowable on my account only through reference to the ways in which they make a difference, this is an interesting epistemological claim. But this makes character traits no different from any other sort of entity postulated to explain various phenomena. Indeed, one recent line of attack against virtue ethics in general argues that there are no character traits (or we have no reason to believe in them) because the explanatory function is better served through reference to the agent's environment than through reference to any internal stable traits of the agent. The advantage that O'Neill grants to the purely ascriptive account seems quite mysterious to me. Kant, of course quite famously argues for the opacity of the will. But he is clearly committed to the metaphysical view that there is a determining motive of the will which either does or does not conform to the categorical imperative – whether we are in a position to know its true nature or not. O'Neill's point is that the effects underdetermine the ascription that we are entitled to make if we have a robust account of the virtues. But recall that the

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

account I offer is an account of virtue across the board – when it comes to specific, particular virtues I am quite willing to argue that they will require certain psychological states – benevolence, for example, does require that one care about the well-being of others, and be motivated to help them. Virtue *per se*, however, does not require this. So I think that O'Neill misunderstands my position (though this is entirely my fault for not being clear enough). Thus, the metaphysical claim is that, to the extent that there are traits which make a positive difference (in one's actions or affective states), they are virtues (whether one can easily identify positive states or not). Indeed, I am quite willing to grant that we cannot do this, and so may misidentify traits as virtues or vices.

I confess to some puzzlement over some of O'Neill's claims. She writes, for example, 'to claim that a trait is a cause of expressions of feeling, or attitudes, or acts we have to view it as knowable and identifiable', and then: 'If traits are inferred from expressions of feeling, from attitudes and from action, they cannot be causes of action.' Both of these claims, but particularly the second, seem very implausible to me. Certainly traits in biology are often seen as causes. I am left unsure whether the claim she makes is metaphysical or epistemological. Again, I grant that it may be difficult to be certain whether or not someone has a virtue simply based on what I happen to see him do; yet, I feel quite confident that he has traits which explain his behaviour.

With regard to the non-consequentialist account she writes that it has the advantage of allowing 'one to take the view that the relation between *dispositions* on the one hand and expressions and feeling and attitude and acts that express them on the other is not causal'. But simply from lack of clarity about causal consequences one cannot infer acausality. For example, it seems quite implausible to hold that honesty does not cause anything, that it does not make a difference, or dishonesty either. After all, it can figure in a causal explanation of someone's behaviour. For some people, they give correct change because they are honest, because they care about fairness, and so on. Again, O'Neill misconceives my project if she thinks I am denying that for particular virtues, there may be states internal to agency crucial to characterizing those virtues. That I believe to be true.

Professor Skorupski, who broadly agrees with the teleological leanings of the account I propose, disagrees also. His view is that the internalist account proposed simply cannot account for our moral practices of appraisal.

He is quite correct to note that my account might well have a problem with self-governance where:

> Central to this conception is a requirement that has been particularly important in modern moral (and political) thought, and is linked to the idea

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

of conscience. It is that moral agents must be able to tell, for themselves, in concrete settings of choice, what is morally right and wrong: they don't need to be told. (p. 2)

We simply disagree on the above claim if the claim is that any moral agent in a given situation is always able to know what the right thing to do is. A person may be in a position to be more or less justified in making a decision given the information available at the time, etc. The way of interpreting this claim which is incompatible with externalism begs the question in favour of the internalist by simply stating that if someone is justified in believing that *x* is correct (right) then it is true, and he has moral knowledge. That is the whole point of dispute. What I agree with when it comes to self-governance is the idea that crucial to moral decision-making is our capacity for choice, for reflective choice. In many situations (though not all) we are better off following our conscience and relying on our own capacity for reflective choice, than in 'being told' by someone else what to do. There are hazards in relying on moral testimony, even of putative experts.

Skorupski's most compelling point is made through the Planet X thought experiment – which is a very nice one, and really made me mentally squirm. But now I would like to make Skorupski squirm a bit. I think he has his own version of the Planet X problem. Suppose the people on Planet X can see our own planet – they can see that whenever Sally pets her kitten, little babies on their own planet are cut with razor blades and suffer horrible agonies. I think that they would regard such acts with the utmost hostility; they would beg and pray for Sally not to pet her kitten, and curse her when she did. I cannot believe that, knowing the bad effects, they would view this in a kindly light. Further, Sally, if by some miracle she was able to discover the connection between her petting the kitten and the far-off suffering of babies, would stop petting the kitten and would feel a great deal of regret and unhappiness at her earlier behaviour. The issue of whether, in her ignorance, we would *blame* her is separate. I suspect we would not, since there would be no point.

Thus, while I believe that Skorupski's thought experiment has a lot of rhetorical punch, I also think that in the end it does not discredit the account I want to develop. We just end up trading rhetorical punches. The account I offered in *Uneasy Virtue* attempted to be sensitive to the sort of issue Skorupski raises with this thought experiment – by noting that what counts as a virtue will be 'context sensitive'. Thus, blind charity might be a virtue in one context and not at all in another; how we determine relevant contexts is, I admit, a very tricky issue – but not something tricky just for objective consequentialists.

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

Skorupski's account is challenging, however, in that he offers an alternative 'internalist' view which he believes can account for the arguments and considerations I raise in favour of externalism. Basically, he holds that a character trait is a virtue only if 'it is a disposition to *respond to* some class of warrantable reasons when they become, in certain contexts, *morally salient* (or: to act from these morally salient reasons non-accidentally)', and then: 'A warrantable reason is *morally salient* in a context when failure by a moral agent in that context to act in accordance with it is (absent extenuation) blameworthy.' Thus, Skorupski agrees with how I use cases to show that an agent can have virtue even though failing to recognize the import of the reasons he is acting on. The Huckleberry Finn case illustrates this. He disagrees, however, in that he feels that this does not sustain an externalist account. What he has done is to show an alternative which can handle some of the cases I present. He views this approach as internalist because it makes use of an internalist account of blame: 'I can and should accept guilt only if *I could have seen that* what I did was wrong' (p. 16, this issue). Like Huck, one need not actually have seen. On my view, a virtue needs result in good systematically, on his view, not. However, on his view when we come to recognize a trait's bad effects, or rather, when those bad effects become recognizable whether they are in fact recognized or not, the trait in question becomes a vice. To me this seems odd – a more natural take on this is to hold that we were mistaken. And, indeed, people do like to think of predecessors who lacked relevant information as mistaken about virtue. Skorupski's account would be incompatible with this.

Professor Slote's comments are very generous. He grants that there may well be some virtues based on ignorance, or at least involving ignorance in some important way – traits such as modesty, for example. However, he suggests that we may admire such traits 'because of the motivation that lies behind the ignorance and the modesty' (p. 24, this issue).

But my view is that there may be no particular motive at all characteristic of a trait like modesty. When one acts modestly one might want to: (i) tell the truth; (ii) be worried about appearing vain; (iii) or actually think one is overestimating when one is underestimating, and thus be motivated by a desire to present oneself well. So I do not think that what we value is some underlying motive (or, at least, we do not value that *alone*). However, I am willing to grant that in the case of other particular virtues, we may well value the motive quite a bit. For example, if we were to do an analysis of 'generosity' I think it quite likely that this virtue requires the agent actually to care about the well-being of others. One would not be generous without this. But this claim is quite distinct from the claim that a certain sort

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

of motive is required for virtue across the board – and that is what I deny in *Uneasy Virtue*. However, I am willing to concede that there may be some non-consequentialist accounts that could accommodate the virtues of ignorance. Slote is correct about that. And, indeed, his own agent-based account might be one of those. Like Skorupski, then, he presents an alternative internalist account, one that can handle the data that I present at the start of the book.

Slote is very sympathetic to my criticisms of the classical view, and the use of cases to show that such a view places too many psychological requirements on virtue. He disagrees, however, with the consequentialist nature of my positive account, favouring instead an account which holds that certain motives are intrinsically valuable – an agent-based view which holds that what is fundamental to moral evaluation are the internal states of the agent.

Slote, on the other hand, wants to defend virtue ethics from the sort of virtue theory that I propose. Strictly speaking, my account of virtue is not at all incompatible with virtue ethics – indeed, one could use it as part of a kind of virtue consequentialism which in turn is another variety, along with rule consequentialism, of indirect consequentialism.[3] However, my account of virtue is incompatible with a view one often sees associated with virtue ethics, and that is the view that virtues are intrinsically valuable. Again, however, it is worth noting that virtue ethicists do not have a lock on this view either. One could be a consequentialist and hold the virtues to be intrinsically valuable.[4]

Slote correctly points out that in its extreme form the view I advocate has counter-intuitive results, and the Mutor case illustrates this. I myself am not comfortable at all with this case, and I believe that when a theory leads to something counter-intuitive the burden is on the theory to explain, or give some kind of error theory regarding those intuitions. Having said that, I also think that we cannot simply go by intuitions in doing moral theory. Then we simply come up with a theory which 'maps' out our intuitions accurately, but has no prescriptive force, and no reformatory aims. Balancing these considerations is a delicate task. However, in *Uneasy Virtue* I did try to provide an error theory that I believe is not incompatible with our views of human goodness. *Given facts about human nature* virtue, in us, does largely involve benevolent motives simply because given how we are wired and facts about the world we get the correlation with positive outcomes. And this correlation has a powerful influence on our intuitions. But alien beings

[3] I discuss this possibility on p. 71 of *Uneasy Virtue*.

[4] See for example Thomas Hurka's account of virtue in *Virtue, Vice, and Value* (New York, 2001).

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press

may be quite different; other rational beings may be wired differently and have evolved in different settings – thus, what goes for them may be different. That is precisely why the case is spelled out in terms of Mutors or aliens, non-human beings.

Slote has pushed me to think, however, in terms of other alternatives to the ones I present in *Uneasy Virtue*. There is no doubt that largely because of his efforts there is a lot more to choose from in the virtue ethics literature today.

Again, I want to thank each of my commentators for their very astute comments and criticisms. Their comments have made me think of new options and alternatives to the view that I propose, and have forced me to clarify my thinking on the issue of what is involved in developing an externalist account of moral virtue. My future research goals have been greatly benefited.

julia.driver@dartmouth.edu

https://doi.org/10.1017/S0953820803001043 Published online by Cambridge University Press